SOME CONTRIBUTIONS TO SUBSET SELECTION PROCEDURES

by

Wing-Yue Wong
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #454

May 1976

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

INTRODUCTION

The classical tests of homogeneity are inadequate in many practical situations in which the experimenter has to make a decision regarding k populations. This inadequacy is not in the development of these tests but rather in the basic formulation itself which is not designed to answer many questions which are of real interest to the experimenter. Bahadur [4], Mosteller [66] and Paulson [71] were among the first research workers to recognize the inadequacy of such tests and to formulate the problem as a multiple decision problem concerned with the ranking and selection of k populations.

In the two decades since these early papers, ranking and selection problems have become an active area of statistical research. In the theory of selection and ranking procedures, there are two basic approaches to the problem. One is called the 'indifference zone' approach and the other is called the 'subset selection' approach. In the former approach, the experimenter is allowed to select, say, one population which is guaranteed to be of interest to him with a fixed probability whenever the unknown parameters lie outside some subset, or zone of indifference, of the entire parameter space. This approach is due to Bechhofer [10].

Other contributions to this approach can be found in Bechhofer and Sobel [14], Bechhofer, Dunnett and Sobel [11], Sobel and Huyett [90]. A quite adequate bibliography may be found in Santner [83] and Bechhofer, Kiefer and Sobel [13].

In contrast to the indifference zone approach, Gupta [32], [35] proposed a formulation in which the experimenter obtains a subset of the k populations for which there is a fixed minimum probability over the whole parameter space that the population of interest is included. This approach is called the subset selection approach. Some recent contributors in the area of subset selection include: Gnanadesikan [26], Gnanadesikan and Gupta [27], Gupta and Studden [49], Gupta and Panchapakesan [46], Gupta and Santner [47]. D. Y. Huang [52] and W. T. Huang [53].

Nearly all of the work in sequential and multistage selection and ranking procedures has been through the indifference approach. Very little work has been done for the multistage subset selection approach. Some sequential subset selection procedures have been proposed and studied by Barron and Gupta [8] and Huang [53].

Some optimum theory results have been developed by Bahadur [4], Bahadur and Goodman [5], Lehmann [63] and Eaton [24]. Contributions toward optimum properties of subset selection procedures have also been made by Goel and Rubin [30], Govindarajulu and Harvey [31], Gupta [33], Gupta and Deely [36], Lehmann [61], Robbins [82], Seal [84], [85], [86] and Studden [93].

The main purpose of this thesis is to propose and study the subset selection approach for some new problems and make contributions.

Chapter I deals with some selection and ranking procedures for the largest unknown mean of k normal populations with unequal variances. The procedures are based on unequal number of observations from the given k normal populations. In Section 1.2, a single sample procedure is proposed and studied under the assumption that the variances are all known. When the variances are unknown to the experimenter, the problem is more difficult than the one above. In this case, a subset selection is proposed and investigated.
In Section 1.3, selection procedures for treatments better than a standard or control are discussed. An indifference zone approach to the problem of selecting the populations with the t-largest unknown means is studied in Section 1.4. In Section 1.5, a test of homogeneity is proposed which is based on the range of sample means.

In Chapter II, we propose and study a nonparametric subset selection procedure based on U-statistics for selecting the largest of the k location parameters. Again, the procedure is based on unequal number of observations from each of the k population. The asymptotic results for the infimum of the probability of a correct selection and the supremum of the expected subset size are given in Section 2.3. It should be pointed out that the procedure proposed in Section 2.3 is based on the complete samples. In Section 2.4, we consider the case of selecting a subset which contains the population with the smallest scale parameter. In Section 2.5, the situation when the samples are trimmed, is studied.

Chapter III discusses some subset selection procedures for Poisson processes. In Section 3.2, three different kind of sampling rules are considered. Some properties of the proposed selection rules are discussed. Section 3.3 deals with the analogous problem of selecting the population with the smallest parameter. In Section 3.4, applications to binomial and multinomial selection problems are discussed.

Chapter IV, deals with a class of selection rules for finite schemes. The parameter space is partially ordered by means of majorization and the proposed selection rules are based on Schur functions. Some properties of the proposed class of selection procedures are discussed in Section 4.3. In Section 4.4, we discuss the procedures when the finite schemes are reduced to binomial case. An upper bound for the expected subset size for the procedure is given in Section 4.5. In Section 4.6, selection of treatments better than a standard or control is discussed. An application to testing the homogeneity of k finite schemes are given in Section 4.7.

Chapter V discusses some subset selection procedures for a negative multinomial distribution. Some properties of the proposed procedures are studied in Section 5.2 and Section 5.3. An inverse sampling rule for selecting the cell with largest cell-probability from a multinomial distribution is discussed in Section 5.4.

CHAPTER I

SUBSET SELECTION PROCEDURES FOR THE MEANS OF NORMAL POPULATIONS WITH UNEQUAL VARIANCES: UNEQUAL SAMPLE SIZES CASE

1.1 INTRODUCTION

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k independent normal populations with unknown means $\mu_1, \mu_2, \ldots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectively. Our goal is to select a nonempty subset of the k populations containing the population with the largest mean. In most of the earlier work (see for example Gupta [35]) it is assumed that either the number of observations from each population is the same or all the populations have a common variance. Very little work has been done in the case of unequal sample sizes and different variances. Sitek [88] proposed a procedure for the normal means; however, her result was shown to be in error by Dudewicz [23]. Recently Gupta and W. T. Huang [42] and Gupta and D. Y. Huang [40] proposed some subset selection procedures for selecting a subset of the unknown normal means. However, all the works mentioned above are based on the assumption that the given k normal populations have a common variance. For the case of unequal variances, Dudewicz [22] proposed a two-sample procedure for the normal means problem. His procedure is based on a linear combination of the first stage sample mean and the second sample mean. In this chapter a procedure based on the overall sample means is proposed and studied.

1.2 SELECTING THE NORMAL POPULATION WITH THE LARGEST MEAN

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k independent normal populations with unknown means $\mu_1, \mu_2, \ldots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectively. The ordered μ_i are denoted by $\mu_{[1]} \le \mu_{[2]} \le \cdots \le \mu_{[k]}$. Here we assume that there is no prior knowledge of the correct pairing of the ordered and unordered μ_i's. Let $X_{i1}, \ldots, X_{in_i}$ be n_i independent observations from population π_i, $i = 1,2,\ldots,k$. Based on these observations, our goal is to select a nonempty subset of the k populations so as to include the population associated with $\mu_{[k]}$. A correct selection (CS) is the selection of any subset containing the population associated with $\mu_{[k]}$. The object is to define a (non-trivial) procedure R so that $P(CS|R)$, the probability of a correct selection, is at least a preassigned number $P^*(\frac{1}{k} < P^* < 1)$ and which has some desirable properties. We shall refer to this requirement as the P*-condition. We shall discuss the two cases: (a) $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ unequal but known, and (b) $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ unequal and unknown.

Case (a): $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ unequal but known.

Let $X_{i1}, \ldots, X_{in_i}$ be n_i independent random samples drawn from populations π_i, $i = 1,\ldots,k$. Let $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ denote the sample mean. We define the following rule R_1 based on these sample means.

R_1: Select the population π_i if and only if

$$\bar{X}_i \geq \max_{1 \leq j \leq k} \left(\bar{X}_j - c_1 \sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}\right) \tag{1.2.1}$$

where $c_1 = c_1(k, P^*, n_1, \ldots, n_k; \sigma_1^2, \ldots, \sigma_k^2)$ is the smallest nonnegative number chosen so as to satisfy the P^*-condition.

Let $\bar{X}_{(i)}$, $n_{(i)}$ and $\sigma_{(i)}^2$ be the sample mean, sample size and variance associated with the population $\pi_{(i)}$ with mean $\mu_{[i]}$, $i = 1, 2, \ldots, k$. It should be pointed out that $\bar{X}_{(i)}$, $n_{(i)}$ and $\sigma_{(i)}^2$ are all unknown. For the evaluation of the infimum of $P(CS|R_1)$, we need a lemma due to Slepian (see Gupta [33]) (stated below without proof).

<u>Lemma 1.2.1</u>. Let $\underset{\sim}{X} = (X_1, \ldots, X_m)$ and $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ be two m-variate normal random vectors with zero mean vector, unit variance and correlation matrices (ρ_{ij}) and (κ_{ij}), respectively. If $\rho_{ij} \geq \kappa_{ij}$ for all $i, j = 1, \ldots, m$, then $P_r(\underset{\sim}{X} \leq \underset{\sim}{a}) \geq P_r(\underset{\sim}{Y} \leq \underset{\sim}{a})$ for all points $\underset{\sim}{a} = (a_1, \ldots, a_m)$ of the real m-Euclidean space, and $\underset{\sim}{X} \leq \underset{\sim}{a}$ means $X_i \leq a_i$ for $i = 1, \ldots, m$.

Let Φ and φ denote the cdf and pdf of a standard normal variate. Now we prove the following theorem regarding the infimum of $P(CS|R_1)$.

<u>Theorem 1.2.1</u>. For the rule R_1 defined in (1.2.1),

$$\min_{n_1, \ldots, n_k} \min_{\sigma_1^2, \ldots, \sigma_k^2} \inf_{\Omega_1} P(CS|R_1) = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \Phi\left(\frac{c_1 - \alpha_j x}{\sqrt{1 - \alpha_j^2}}\right) d\Phi(x) \tag{1.2.2}$$

where $\Omega_1 = \{\underline{\mu}: \underline{\mu} = (\mu_1,\ldots,\mu_k), -\infty < \mu_i < \infty, i = 1,2,\ldots,k\}$ and

$$(1.2.3) \qquad \alpha_j = \left(1 + \frac{\sigma^2_{[k-j+1]}}{\sigma^2_{[1]}} \cdot \frac{n_{[k]}}{n_{[j]}}\right)^{-\frac{1}{2}}, \; j = 1,\ldots,k-1;$$

$\sigma_{[1]} \leq \sigma_{[2]} \leq \cdots \leq \sigma_{[k]}$ and $n_{[1]} \leq n_{[2]} \leq \cdots \leq n_{[k]}$ denote the ordered values of the given sets $\{\sigma_1,\ldots,\sigma_k\}$ and $\{n_1,\ldots,n_k\}$ respectively.

Proof.

$$P(CS|R_1) = P_r\left(\bar{X}_{(k)} \geq \max_{1\leq j\leq k-1}\left(\bar{X}_{(j)} - c_1\sqrt{\frac{\sigma^2_{(k)}}{n_{(k)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}}\right)\right)$$

$$= P_r\left(\frac{\bar{X}_{(j)} - \bar{X}_{(k)}}{\sqrt{\frac{\sigma^2_{(j)}}{n_{(j)}} + \frac{\sigma^2_{(k)}}{n_{(k)}}}} \leq c_1, \quad j = 1,\ldots,k-1\right)$$

$$(1.2.4) \qquad = P_r\left(Z_{jk} \leq c_1 + \frac{\mu_{[k]} - \mu_{[j]}}{\sqrt{\frac{\sigma^2_{(k)}}{n_{(k)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}}}, \quad j = 1,\ldots,k-1\right)$$

where for $j = 1,\ldots,k-1$, Z_{jk} is given by

$$(1.2.5) \qquad Z_{jk} = \frac{\bar{X}_{(j)} - \bar{X}_{(k)} - \mu_{[j]} + \mu_{[k]}}{\sqrt{\frac{\sigma^2_{(j)}}{n_{(j)}} + \frac{\sigma^2_{(k)}}{n_{(k)}}}} .$$

Thus $(Z_{1k}, Z_{2k}, \ldots, Z_{k-1\,k})$ is a $(k-1)$-variate normal with zero mean vector, unit variance and correlation matrix

$$(1.2.6)\qquad \rho_{ij}^{(k)} = \{(1 + \frac{\sigma_{(i)}^2}{\sigma_{(k)}^2} \cdot \frac{n_{(k)}}{n_{(i)}})\ (1 + \frac{\sigma_{(j)}^2}{\sigma_{(k)}^2} \cdot \frac{n_{(k)}}{n_{(j)}})\}^{-\frac{1}{2}},$$

$$i,j = 1,\ldots,k-1,\ i \neq j.$$

We see from (1.2.4) that for any given association between $\{n_1,\ldots,n_k\}$, $\{\sigma_1^2,\ldots,\sigma_k^2\}$ and $\{n_{(1)},\ldots,n_{(k)}\}$, $\{\sigma_{(1)}^2,\ldots,\sigma_{(k)}^2\}$ respectively, the infimum of $P(CS|R_1)$ will be attained when $\mu_{[1]} =\ldots= \mu_{[k]}$. Thus the infimum we seek in (1.2.2) is reduced to

$$(1.2.7)\qquad \min_{n_1,\ldots,n_k}\ \min_{\sigma_1,\ldots,\sigma_k}\ P_r(Z_{jk} \leq c_1,\ \ j = 1,\ldots,k-1)$$

where Z_{jk} are defined by (1.2.5). For any fixed $\sigma_{(1)},\ldots,\sigma_{(k)}$, and $1 \leq \ell \leq k$, if we let

$$\kappa_{ij}^{(\ell)}(r,s) = \{(1 + \frac{\sigma_{(r)}^2}{\sigma_{(s)}^2}\ \frac{n_{[\ell]}}{n_{[i]}})(1 + \frac{\sigma_{(r)}^2}{\sigma_{(s)}^2} \cdot \frac{n_{[\ell]}}{n_{[j]}})\}^{-\frac{1}{2}},$$

$$(1.2.8)\qquad i,j = 1,\ldots,k;\ i,j \neq \ell,\ i \neq j,$$

$$\kappa_{ii}^{(\ell)}(r,s) = 1 \qquad ,\ i = 1,\ldots,k,\ i \neq \ell.$$

where $n_{[1]} \leq\cdots\leq n_{[k]}$ denote the ordered values of a given set $n_1,\ldots,n_k$, it follows from the fact that $n_{[\ell]} \leq n_{[k]}$ for all $i = 1,\ldots,k-1$, we have for any (r,s),

$$(1.2.9)\qquad \kappa_{i(j+1)}^{(\ell)}(r,s) \geq \kappa_{ij}^{(\ell)}(r,s) \geq \kappa_{ij}^{(k)}(r,s) \text{ for all } i,j=1,\ldots,k-1,$$

$i \neq \ell,k$. By Lemma 1.2.1, this implies that

$$(1.2.10)\quad \min_{n_1,\ldots,n_k}\ \min_{\sigma_1,\ldots,\sigma_k} P_r(Z_{jk} \le c_1, j=1,\ldots,k-1) = \min_{\sigma_1,\ldots,\sigma_k} P_r(Y_{jk} \le c_1,\ j = 1,\ldots,k-1)$$

where $(Y_{1k},\ldots,Y_{k-1\ k})$ represents a (k-1)-variate normal with zero mean vector, unit variance and correlation matrix $(\zeta_{ij}^{(k)})$, where

$$(1.2.11)\quad \zeta_{ij}^{(k)} = \{(1 + \frac{\sigma_{(i)}^2}{\sigma_{(k)}^2} \frac{n_{[k]}}{n_{[i]}})\ (1 + \frac{\sigma_{(j)}^2}{\sigma_{(k)}^2} \cdot \frac{n_{[k]}}{n_{[j]}})\}^{-\frac{1}{2}},$$

$$i,j = 1,\ldots,k-1,\ i \neq j.$$

Similarly if we let

$$\tilde{\zeta}_{ij}^{(\ell)} = \{(1 + \frac{\sigma_{[k-i+1]}^2}{\sigma_{[\ell]}^2} \cdot \frac{n_{[k]}}{n_{[i]}})\ (1 + \frac{\sigma_{[k-j+1]}^2}{\sigma_{[\ell]}^2} \cdot \frac{n_{[k]}}{n_{[j]}})\}^{-\frac{1}{2}} \text{ for}$$

$$i,j = 1,\ldots,k,\ i,j \neq \ell,\ i \neq j$$

$$\tilde{\zeta}_{ii}^{(\ell)} = 1 \qquad \text{for } i = 1,\ldots,k,\ \ i \neq \ell,$$

then by using an analogous argument, it follows that

$$(1.2.12)\quad \min_{\sigma_1,\ldots,\sigma_k} P_r(Y_{jk} \le c_1,\ j = 1,\ldots,k-1) = P_r(Y_{jk}^* \le c_1,\ j=1,\ldots,k-1)$$

where $(Y_{1k}^*,\ldots,Y_{k-1\ k}^*)$ is a (k-1)-variate normal with zero mean vector, unit variance and correlation matrix $(\tilde{\zeta}_{ij}^{(1)})$. Let

$$(1.2.13)\quad \alpha_i = (1 + \frac{\sigma_{[k-i+1]}^2}{\sigma_{[1]}^2} \cdot \frac{n_{[k]}}{n_{[i]}})^{-\frac{1}{2}},\quad i = 1,\ldots,k-1,$$

It is well-known that $Y^*_{1k},\ldots,Y^*_{k-1\,k}$ can be generated from k independent standard normal variates $Z_1,Z_2,\ldots,Z_k$ by the transformation

$$(1.2.14)\qquad Y^*_{ik} = (1-\alpha_i^2)^{\frac{1}{2}} Z_i + \alpha_i Z_k, \qquad i = 1,\ldots,k-1.$$

Hence the right hand side of (1.2.12) can be rewritten as follows:

$$(1.2.15)\qquad \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \Phi\left(\frac{c_1-\alpha_j x}{\sqrt{1-\alpha_j^2}}\right) d\Phi(x).$$

This completes the proof of the theorem.

It should be pointed out that when $\sigma_{[1]} = \ldots = \sigma_{[k]} = \sigma$, say, the expression (1.2.15) is independent of σ, this reduces to the result obtained by Gupta and Huang [40].

Consistent with the basic probability requirement, we would like the size of the selected subset to be small. Now, S, the size of the selected subset is a random variable which takes values 1,2,...,k. Hence one can use as a criterion of the efficiency of the procedure R_1, the expected value of the size of the selected subset.

$$(1.2.16)\qquad E(S|R_1) = \sum_{i=1}^{k} P(\pi_{(i)} \text{ is selected } |R_1)$$

$$= \sum_{i=1}^{k} P_r\left(\frac{\bar{X}_{(j)} - \bar{X}_{(i)}}{\sqrt{\frac{\sigma^2_{(j)}}{n_{(j)}} + \frac{\sigma^2_{(i)}}{n_{(i)}}}} \le c_1 \qquad j = 1,\ldots,k,\ j \ne i\right)$$

<u>Theorem 1.2.2</u>. For the rule R_1,

$$(1.2.17) \qquad \max_{n_1,\ldots,n_k} \max_{\sigma_1,\ldots,\sigma_k} \sup_{\Omega_1} E_{\Omega_1}(S|R_1) \le k\,\Phi(c_1).$$

Proof. Arguing along the lines of the proof of Theorem 2.2 of Gupta and Huang [40], we have

$$E_{\Omega_1}(S|R_1) = \sum_{i=1}^{k} P(\pi_{(i)} \text{ is selected } |R_1)$$

$$= \sum_{i=1}^{k} P_r\{ \max_{\substack{1\le j\le k \\ j\ne i}} \left(\frac{\bar{X}_{(j)} - \bar{X}_{(i)}}{\sqrt{\frac{\sigma^2_{(j)}}{n_{(j)}} + \frac{\sigma^2_{(i)}}{n_{(i)}}}}\right) \le c_1\}$$

$$\le \frac{1}{k-1} \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\ne i}}^{k} P_r\left(\frac{\bar{X}_{(j)} - \bar{X}_{(i)}}{\sqrt{\frac{\sigma^2_{(j)}}{n_{(j)}} + \frac{\sigma^2_{(i)}}{n_{(i)}}}} \le c_1\right)$$

$$= \frac{1}{k-1} \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\ne i}}^{k} \Phi\left(c_1 + (\mu_{[j]} - \mu_{[i]})\left(\frac{\sigma^2_{(j)}}{n_{(j)}} + \frac{\sigma^2_{(i)}}{n_{(i)}}\right)^{-\frac{1}{2}}\right)$$

$$(1.2.18) \qquad = \frac{1}{k-1} Q \quad \text{(say)}.$$

Now we show that Q attains its supremum when $\mu_{[1]} = \ldots = \mu_{[k]}$. To this end, we consider the configuration

$$(1.2.19) \qquad \mu_{[1]} = \ldots = \mu_{[m]} = \mu \le \mu_{[m+1]} \le \cdots \le \mu_{[k]}, \quad 1 \le m \le k-1.$$

For this configuration, Q can be rewritten as follows:

$$(1.2.20) \quad Q = \sum_{i=1}^{m} \{(m-1)\Phi(c_1) + \sum_{j=m+1}^{k} \Phi\left(c_1+(\mu-\mu_{[j]})\left(\frac{\sigma^2_{(i)}}{n_{(i)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}\right)^{-\frac{1}{2}}\right)\}$$

$$+ \sum_{i=m+1}^{k} \{ \sum_{j=1}^{m} \Phi\left(c_1+(\mu_{[i]}-\mu)\left(\frac{\sigma^2_{(i)}}{n_{(i)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}\right)^{-\frac{1}{2}}\right) +$$

$$+ \sum_{\substack{j=m+1 \\ j\neq i}}^{k} \Phi(c_1+(\mu_{[i]}-\mu_{[j]})(\frac{\sigma_{(i)}^2}{n_{(i)}} + \frac{\sigma_{(j)}^2}{n_{(j)}})^{-\frac{1}{2}})\}.$$

Interchange the labels i and j in the sum $\sum_{i=1}^{m} \sum_{j=m+1}^{k}$ and then differentiating with respect to μ and grouping the terms, we have

$$\frac{\partial Q}{\partial \mu} = \sum_{i=m+1}^{k} \sum_{j=1}^{m} (\frac{\sigma_{(i)}^2}{n_{(i)}} + \frac{\sigma_{(j)}^2}{n_{(j)}})\{\varphi(c_1+(\mu-\mu_{[i]})(\frac{\sigma_{(i)}^2}{n_{(i)}} + \frac{\sigma_{(j)}^2}{n_{(j)}})^{-\frac{1}{2}})$$

$$- \varphi(c_1+(\mu_{[i]}-\mu)(\frac{\sigma_{(i)}^2}{n_{(i)}} + \frac{\sigma_{(j)}^2}{n_{(j)}})^{-\frac{1}{2}})\}$$

$$\geq 0$$

where $\varphi(\cdot)$ represents the pdf of the standard normal variate. Thus, by successive application of the above result, with $m = 1,\ldots,k-1$, we see that the supremum of Q over Ω_1 is attained when $\mu_{[1]} =\ldots= \mu_{[k]}$ and this gives

(1.2.21) $$\sup_{\Omega_1} E(S|R_1) \leq \frac{1}{k-1}(k-1)k\Phi(c_1) = k\,\Phi(c_1).$$

Since this is true for all possible association between $\{n_1,\ldots,n_k\}$ and $\{n_{(1)},\ldots,n_{(k)}\}$, and between $\{\sigma_1,\ldots,\sigma_k\}$ and $\{\sigma_{(1)},\ldots,\sigma_{(k)}\}$, the result (1.2.17) follows.

Remark 1.2.1: For k = 2,

(1.2.22) $$\int_{-\infty}^{\infty} \Phi(\frac{c_1-\alpha_1 x}{\sqrt{1-\alpha_1^2}})d\Phi(x) = \Phi(c_1).$$

It follows that the constant c_1 obtained to satisfy the P*-condition is given by $\Phi(c_1) = P^*$. Thus in this case the upper bound of

$E(S|R_1)$ is $2P^*$, which is the exact upper bound in the case of equal sample size and equal known variance.

Case (b) $\sigma_1^2,\ldots,\sigma_k^2$ unknown.

As pointed out earlier, this case presents more difficulty than the case in which $\sigma_1^2,\ldots,\sigma_k^2$ are assumed known. We propose and investigate a selection procedure for this case as described below. For this problem it is necessary

to require that $n_i \geq 2$ for all n_i where n_i is the total number of independent observations from π_i, $i = 1,\ldots,k$. We now define the selection procedure as follows:

(i) Take a first sample of size $n_0 (\geq 2)$ independent observations from each population π_i, $i = 1,\ldots,k$.

(ii) Calculate for $i = 1,\ldots,k$ the sample means and variances based on these n_0 observations,

(1.2.23)
$$\tilde{X}_i = \frac{1}{n_0}\sum_{j=1}^{n_0} X_{ij}$$
$$S_i^2 = \frac{1}{n_0-1}\sum_{j=1}^{n_0} (X_{ij} - \tilde{X}_i)^2$$

where X_{ij} represents the jth observation from π_i, $i = 1,\ldots,k$, $j = 1,\ldots,n_0$.

Let

(1.2.24)
$$V^2 = \max_{1\leq i\leq k} \frac{S_i^2}{n_i} .$$

(iii) Take the additional $n_i - n_0$ independent observations from π_i, $i = 1,\ldots,k$. Calculate the sample means based on the total n_i observations for π_i,

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad i = 1,\ldots,k,$$

and retain the population π_i in the selected subset if and only if

$$(1.2.25) \quad R_2: \quad \bar{X}_i \geq \max_{1 \leq j \leq k} \bar{X}_j - c_2 V$$

where $c_2 = c_2(k,P^*,n_1,\ldots,n_k,n_0) > 0$ is the smallest positive number such that the P*-condition holds.

To obtain a lower bound for the infimum of $P(CS|R_2)$, we see that

$$P(CS|R_2) = P_r(\bar{X}_{(k)} \geq \max_{1 \leq i \leq k-1} \bar{X}_{(i)} - c_2 V)$$

$$(1.2.26) \qquad = P_r(Z_{ik} \leq (c_2 V + \mu_{[k]} - \mu_{[i]})(\frac{\sigma^2_{(k)}}{n_{(k)}} + \frac{\sigma^2_{(i)}}{n_{(i)}})^{-\frac{1}{2}}, \quad i = 1,\ldots,k-1)$$

where Z_{ik} is given by (1.2.5). It follows from (1.2.26) that

$$(1.2.27) \quad P(CS|R_2) \geq P_r(Z_{ik} \leq c_2 V(\frac{\sigma^2_{(k)}}{n_{(k)}} + \frac{\sigma^2_{(i)}}{n_{(i)}})^{-\frac{1}{2}}, \quad i = 1,\ldots,k-1)$$

$$\geq P_r(Z_{ik} \leq c_2(\frac{\sigma^2_{(k)}}{S^2_{(k)}} + \frac{\sigma^2_{(i)}}{S^2_{(i)}})^{-\frac{1}{2}}, \quad i = 1,\ldots,k-1)$$

Denote the right hand member of (1.2.27) by T. Since the entities of the correlation matrix of $(Z_{1k},\ldots,Z_{k-1\,k})$ are all nonnegative, it follows that given $S^2_{(1)},\ldots,S^2_{(k)}$, T is minimized when $\sigma_{(k)}$ approaches zero. Or more precisely,

$$(1.2.27)\qquad T \geq P_r(Z^*_{ik} \leq c_2(\frac{\sigma^2_{(k)}}{S^2_{(k)}} + \frac{\sigma^2_{(i)}}{S^2_{(i)}})^{-\frac{1}{2}}, \quad i = 1,\ldots,k-1)$$

where $Z^*_{1k},\ldots,Z^*_{k-1\,k}$ are iid standard normal variates, and are independent of $S_{(1)},\ldots,S_{(k)}$. It follows that

$$(1.2.28)\qquad T \geq \int_0^\infty \{\int_0^\infty \Phi(\frac{c_2}{\sqrt{(n_0-1)(\frac{1}{x}+\frac{1}{y})}})dG(x)\}^{k-1}dG(y)$$

where $G(\cdot)$ denote the cdf of chi-square random variate with (n_0-1) degree of freedom. Thus infimum of $P(CS|R_2)$ is at least P^* if c_2 is determined by the equation

$$(1.2.29)\qquad \int_0^\infty\{\int_0^\infty\Phi(\frac{c_2}{\sqrt{(n_0-1)(\frac{1}{x}+\frac{1}{y})}})dG(x)\}^{k-1}dG(y) = P^*.$$

Thus we have shown the following theorem.

<u>Theorem 1.2.3</u>. If c_2 is defined by (1.2.29), then

$$\min_{n_1,\ldots,n_k}\ \inf_{\Omega_2} P(CS|R_2) \geq P^*$$

where $\Omega_2 = \{(\mu_1,\ldots,\mu_k;\ \sigma_1,\ldots,\sigma_k):\ -\infty < \mu_i < \infty,\ \sigma_i > 0,\ i = 1,\ldots,k\}$.

Next we consider the expected subset size $E(S|R_2)$. It is given by

$$E(S|R_2) = \sum_{i=1}^k P(\pi_{(i)} \text{ is selected } |R_2)$$

$$= \sum_{i=1}^k P_r(\max_{\substack{1\leq j\leq k\\ j\neq i}} (\bar{X}_{(j)} - \bar{X}_{(i)}) \leq c_2 v).$$

Using the similar argument as in the proof of Theorem 1.2.2, it is easy to see that

$$(1.2.31)\quad E(S|R_2) \leq \frac{1}{k-1} \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \int_0^{\infty} \Phi\left(c_2 - \frac{v}{\xi_{ij}^{\frac{1}{2}}} + \frac{\mu_{[i]}-\mu_{[j]}}{\xi_{ij}^{\frac{1}{2}}}\right) dH(v)$$

where $H(\cdot)$ represents the cdf of V, and $\xi_{ij} = \left(\frac{\sigma^2_{(i)}}{n_{(i)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}\right)$, $i,j = 1,\ldots,k$, $i \neq j$. Denote the right hand number of (1.2.31) by $\frac{1}{k-1} \int Q_2\, dH(v)$, and consider the configuration (1.2.19) and suppose $\delta_2 > \sigma_i > \delta_1 > 0$ for all $i = 1,\ldots,k$.

$$(1.2.32)\quad Q_2 = \sum_{i=1}^{m} \Big\{ \sum_{\substack{j=1 \\ j\neq i}}^{m} \Phi\Big(\frac{c_2 v}{\sqrt{\xi_{ij}}}\Big) + \sum_{j=m+1}^{k} \Phi((c_2 v + (\mu-\mu_{[j]}))\xi_{ij}^{-\frac{1}{2}}) \Big\}$$
$$+ \sum_{i=m+1}^{k} \Big\{ \sum_{j=1}^{m} \Phi((c_2 v+(\mu_{[i]}-\mu))\xi_{ij}^{-\frac{1}{2}}) + \sum_{\substack{j=m+1 \\ j\neq i}}^{k} \Phi((c_2 v + (\mu_{[i]}-\mu_{[j]}))\xi_{ij}^{-\frac{1}{2}}) \Big\}.$$

Keeping all parametersbut μ fixed and differentiating Q_2 with respect to μ and interchanging the labels i and j in the sum $\sum_{i=1}^{m} \sum_{j=m+1}^{k}$, we obtain

$$(1.2.33)\quad \frac{\partial Q_2}{\partial \mu} = \sum_{i=1}^{m} \sum_{j=m+1}^{k} \xi_{ij}^{-\frac{1}{2}} \{ \varphi((c_2 v+(\mu-\mu_{[j]}))\xi_{ij}^{-\frac{1}{2}}) - \varphi((c_2 v + (\mu_{[j]}-\mu))\xi_{ij}^{-\frac{1}{2}}) \}$$
$$\geq 0$$

This shows that

$$(1.2.34) \qquad Q_2 \le \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \ne i}}^{k} \Phi\left(\frac{c_2 v}{\sqrt{\xi_{ij}}}\right)$$

$$\le \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \ne i}}^{k} \Phi\left(\frac{\delta_2}{\delta_1} \frac{\max \sqrt{n_0 - 1}\, \frac{s_i}{\sigma_i}}{\sqrt{(n_0-1)\left(\frac{n_{[1]}}{n_i} + \frac{n_{[1]}}{n_j}\right)}}\right)$$

$$= 2 \sum_{1 \le i < j \le k} \Phi\left(\frac{\delta_2}{\delta_1} \frac{\max \sqrt{n_0 - 1}\, \frac{s_i}{\sigma_i}}{\sqrt{(n_0-1)\left(\frac{n_{[1]}}{n_i} + \frac{n_{[1]}}{n_j}\right)}}\right)$$

Let $\mathcal{A}$ represent the cdf of $\max \sqrt{n_0 - 1}\, \frac{s_i}{\sigma_i}$. Combining (1.2.31) and (1.2.34), we obtain the following theorem.

Theorem 1.2.4. Let $\Omega_2(\delta_1, \delta_2) = \{(\mu_1, \ldots, \mu_k; \sigma_1, \ldots, \sigma_k): -\infty < \mu_i < \infty,$ $\delta_2 > \sigma_i > \delta_1 > 0, \ i = 1, \ldots, k\}$.

$$(1.2.35) \qquad \max_{n_1, \ldots, n_k} \sup_{\Omega_2(\delta_1, \delta_2)} E(S|R_2) \le k \frac{1}{\binom{k}{2}} \sum_{1 \le i < j \le k} \int_0^\infty \Phi\left(\frac{\delta_2}{\delta_1} \frac{x}{\sqrt{(n_0-1)\left(\frac{n_{[1]}}{n_{[i]}} + \frac{n_{[1]}}{n_{[j]}}\right)}}\right) d\mathcal{A}(x).$$

1.3 SELECTING A SUBSET WHICH CONTAINS ALL POPULATIONS BETTER THAN A STANDARD

In this section, we discuss a related selection problem.

Let $\pi_0, \pi_1, \ldots, \pi_k$ be k+1 independent normal populations with means $\mu_0, \mu_1, \ldots, \mu_k$ and variances $\sigma_0^2, \sigma_1^2, \ldots, \sigma_k^2$, respectively. It is assumed that $\mu_0, \mu_1, \ldots, \mu_k$ are unknown. The procedure described in this section controls the probability that the selected subset contains all those populations better than the standard ($\mu_i \geq \mu_0$), with the probability of a correct decision to be at least P*. Again, we discuss separately the following cases:

Case A. $\sigma_0^2, \sigma_1^2, \ldots, \sigma_k^2$ are known

Let $\bar{X}_i$ denote the mean based on a sample size n_i independent observations taken from population π_i, $i = 0,1,\ldots,k$. We propose a procedure as follows:

R_3: Retain in the selected subset those and only those populations π_i $(i = 1,\ldots,k)$ for which

$$(1.3.1) \qquad \bar{X}_i \geq \bar{X}_0 - c_3\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_i^2}{n_i}} \; .$$

Let r_1 and r_2 denote the number of populations with $\mu \geq \mu_0$ and $\mu < \mu_0$, respectively, so that $r_1 + r_2 = k$. The probability of a correct decision (CD) is given by

$$(1.3.2) \quad P(CD|R_3) = P_r(\bar{X}_{(i)} \geq \bar{X}_0 - c_3\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_{(i)}^2}{n_{(i)}}},\ i = r_2+1,\ldots,k)$$

$$= P_r(Z_i \leq c_3 + \frac{\mu_{[i]} - \mu_0}{\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_{(i)}^2}{n_{(i)}}}}, \qquad i = r_2+1,\ldots,k)$$

where $(Z_{r_2+1},\ldots,Z_k)$ is a r_1-variate normal with zero mean vector, unit variances and correlation matrix (ρ_{ij}) where

$$\rho_{ij} = \{(1 + \frac{n_0}{n_{(i)}} \frac{\sigma^2_{(i)}}{\sigma^2_0})(1 + \frac{n_0}{n(j)} \frac{\sigma^2_{(j)}}{\sigma^2_0})\}^{-\frac{1}{2}} \qquad i,j = r_2+1,\ldots,k, \; i \neq j$$

(1.3.3)

$$\rho_{ii} = 1$$

By using the transformation (1.2.14), it follows that $P(CD|R_3)$ is bounded below by

$$\int_{-\infty}^{\infty} \prod_{j=1}^{k} \Phi(\frac{c_3-\alpha_j x}{\sqrt{1-\alpha_j^2}})d\Phi(x)$$

where

(1.3.4) $$\alpha_j = (1 + \frac{n_0}{n_j} \frac{\sigma_j^2}{\sigma_0^2})^{-\frac{1}{2}}, \qquad j = 1,\ldots,k.$$

Hence

Theorem 1.3.1. For rule R_3,

(1.3.5) $$P(CD|R_3) \geq \int_{-\infty}^{\infty} \prod_{j=1}^{k} \Phi(\frac{c_3-\alpha_j x}{\sqrt{1-\alpha_j^2}})d\Phi(x)$$

Let S_3 denote the number of populations with means less than μ_0 that are included in the selected subset.

$$\begin{aligned} E(S_3|R_3) &= \sum_{i=1}^{r_2} P(\pi_{(i)} \text{ is selected } |R_3) \\ &= \sum_{i=1}^{r_2} P_r(\bar{X}_{(i)} \geq \bar{X}_0 - c_3\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_{(i)}^2}{n_{(i)}}}) \\ &= \sum_{i=1}^{r_2} P_r(\frac{\bar{X}_0-\bar{X}_{(i)}-\mu_0+\mu_{[i]}}{\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_{(i)}^2}{n_{(i)}}}} \leq c_3 - \frac{\mu_0-\mu_{[i]}}{\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_{(i)}^2}{n_{(i)}}}}) \\ &\leq r_2 \, \Phi(c_3). \end{aligned}$$

Case B. $\sigma_0, \sigma_1, \ldots, \sigma_k$ unknown.

In this case we propose a selection procedure to select those populations whose means are greater than or equal to μ_0. Again we assume that $n_i \geq m \geq 2$ for $i = 0,1,\ldots,k$. As in the case (b) of Section 1.2.

(i) Take a first sample of size m observations from each population π_i, $i = 0,1,\ldots,k$.

(ii) Calculate for $i = 0,1,\ldots,k$ the sample means and variances based on these observations.

$$\tilde{X}_i = \frac{1}{m} \sum_{j=1}^{m} X_{ij}$$

(1.3.6)

$$S_i^2 = \frac{1}{m-1} \sum_{j=1}^{m} (X_{ij} - \tilde{X}_i)^2$$

where X_{ij} represents the jth observation from π_i, $i = 0,1,\ldots,k$.

Let

$$V^2 = \max_{0 \leq i \leq k} \frac{S_i^2}{n_i} . \tag{1.3.7}$$

Take the remaining n_i-m observations from each population π_i and compute the overall sample mean

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \qquad i = 0,1,\ldots,k.$$

Now we propose a selection procedure as follows:

R_4: Include population π_i in the selected subset if and only if

(1.3.8) $$\bar{X}_i \geq \bar{X}_0 - c_4 V$$

where $c_4 = c_4(k,P^*,m,n_0,n_1,\ldots,n_k)$ is the smallest positive number such that the basic P*-condition is satisfied. The probability of a correct decision is given by

(1.3.9) $$P(CD|R_4) = P_r(\bar{X}_0 - \bar{X}_{(i)} \leq c_4 V, \quad i = r_2+1,\ldots,k)$$

where r_2 denote the number of populations whose means are less than μ_0. Now $P(CD|R_4)$ can be expressed as follows:

(1.3.10) $$P(CD|R_4) = P_r(Z_i \leq (c_4 V+(\mu_{[i]}-\mu_0))\xi_i^{-\frac{1}{2}}, \ i = r_2+1,\ldots,k)$$

where

$$\xi_i = \frac{\sigma_0^2}{n_0} + \frac{\sigma_{(i)}^2}{n_{(i)}}$$

(1.3.11)

$$Z_i = (\bar{X}_0 - \bar{X}_{(i)} + \mu_{[i]} - \mu_0)\xi_i^{-\frac{1}{2}}.$$

Using the similar argument as in the proof of Theorem 1.2.2, we obtain

(1.3.12) $$P(CD|R_4) \geq \int_0^\infty \{\int_0^\infty \Phi(\frac{c_4}{\sqrt{(m-1)(\frac{1}{x}+\frac{1}{y})}})dG(x)\}^{r_1} dG(y)$$

where $r_1 = k-r_2$ and $G(\cdot)$ represents the cdf of the chi-square random variate with (m-1) degree of freedom. Hence we have shown the following theorem.

Theorem 1.3.2. For procedure R_4, if c_4 is determined by the equation

$$\int_0^\infty \{\int_0^\infty \Phi(\frac{c_4}{\sqrt{(m-1)(\frac{1}{x}+\frac{1}{y})}})dG(x)\}^k dG(y) = P^*$$

then

$$\min_{n_1,\ldots,n_k} \inf_{\Omega_3} P(CD|R_4) \geq P^*.$$

where $\Omega_3 = \{(\mu_0,\mu_1,\ldots,\mu_k,\sigma_0^2,\sigma_1^2,\ldots,\sigma_k^2)\colon\ -\infty < \mu_i < \infty,\ \sigma_i \geq 0,$ $i = 0,1,\ldots,k\}$.

Let S_4 be the number of populations with means less than μ_0 that are included in the selected subset.

$E(S_4|R_4)$

$$= \sum_{i=1}^{r_2} P_r(\bar{X}_{(i)} > \bar{X}_0 - c_4 V)$$

$$= \sum_{i=1}^{r_2} P_r\left(\frac{\bar{X}_0-\bar{X}_{(i)}-\mu_0+\mu_{[i]}}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_{(i)}^2}{n_{(i)}}}} \leq \frac{1}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_{(i)}^2}{n_{(i)}}}}(c_4 V - \mu_0 + \mu_{[i]})\right)$$

$$\leq \sum_{i=1}^{r_2} P_r\left(\frac{\bar{X}_0-\bar{X}_{(i)}-\mu_0+\mu_{[i]}}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_{(i)}^2}{n_{(i)}}}} \leq \frac{Vc_4}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_{(i)}^2}{n_{(i)}}}} \quad \frac{1}{\sqrt{m-1}}\right).$$

This implies that

$$E(S_4|R_4) \leq r_2 \cdot \int_0^\infty \{\int_0^\infty \Phi(\frac{c_4}{\sqrt{(m-1)(\frac{1}{x}+\frac{1}{y})}})dG(x)\}dG(y).$$

1.4 SELECTING THE POPULATIONS WITH THE t-LARGEST MEANS FROM SEVERAL NORMAL POPULATIONS WITH UNKNOWN VARIANCES - AN INDIFFERENCE ZONE APPROACH

It will be helpful if we first state the problem clearly. An experimenter is given k normal populations π_i with unknown means μ_i and unknown variances σ_i^2, $i = 1,\ldots,k$. Let the ranked μ-values be denoted by $\mu_{[1]} \leq \cdots \leq \mu_{[k]}$. It is assumed that there is no a priori information about the true pairing of π_i with $\pi_{[j]}$ $(i,j = 1,\ldots,k)$. The experimenter is interested in selecting an unordered set of t ($\geq$ 1) "best" populations which are associated with the t-largest means $\mu_{[k-t+1]},\ldots,\mu_{[k]}$. Before the experimentation, he has to specify a pair of constant (δ^*,P^*) where $0 < \delta^* < \infty$, $\frac{1}{\binom{k}{t}} < P^* < 1$, with the intention of achieving the following probability requirement

(1.4.1) $\quad P(CS) \geq P^*$ whenever $\mu_{[k-t+1]} - \mu_{[k-t]} \geq \delta^*$.

In order to solve this problem, we need two stage (samples) procedure of the type proposed by Stein [92]. The steps are as follows:

(i) Take an initial sample of size n_0 from each of the population and compute the separate sample variances

(1.4.2) $$S_i^2 = \frac{1}{n_0-1}\sum_{j=1}^{n_0}(X_{ij} - \tilde{X}_i)^2$$

where $\tilde{X}_i = \frac{1}{n_0}\sum_{j=1}^{n_0} X_{ij}$, $i = 1,\ldots,k$. X_{ij} represents the jth observation from population π_i.

(ii) Take a further sample of size $n_i - n_0$ from π_i, $i = 1,\ldots,k$, where

(1.4.3) $$n_i = \max\{n_0, [(\frac{s_i c}{\delta^*})^2] + 1\}$$

where [x] denote the largest integer less than or equal to x, and $c = c(P^*, n_0, k, t)$ is a positive constant which will be defined in equation (1.4.4).

(iii) Calculate from π_i the overall sample mean $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$, $i = 1,\ldots,k$, and select the t populations which have given rise to the t largest overall sample means.

Remark 1.4.1. The sampling rule in this section is slightly different from the one given in Section 1.2.

Theorem 1.4.1. Let C satisfy the equation

(1.4.4) $$\int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{t} \{\Phi(\frac{c}{\sqrt{(n_0-1)(\frac{1}{x} + \frac{1}{y_i})}}) dG(x)\}^{k-t} dG(y_i) = P^*$$

where $G(\cdot)$ represents the cdf of a chi-square random variate with (n_0-1) degree of freedom, then $P(CS) \geq P^*$ whenever $\mu_{[k-t+1]} - \mu_{[k-t]} \geq \delta^*$.

Proof. Note that

(1.4.5) $$P(CS) = P_r(\min_{k-t+1 \leq i \leq k} \bar{X}_{(i)} \geq \max_{1 \leq j \leq k-t} \bar{X}_{(j)}).$$

Using a theorem of Barr and Rizvi [7], it follows that P(CS) is non-increasing in $\theta_{[j]}$, $j = 1,\ldots,k-t$ and nondecreasing in $\theta_{[i]}$, $i = k-t+1,\ldots,k$. Consider the configuration

(1.4.6) $\theta_{[1]} = \ldots = \theta_{[k-t]} = \theta_{[k-t+1]} - \delta^* = \ldots = \theta_{[k]} - \delta^*$.

Under this configuration, P(CS) can be written as follows

$$\text{(1.4.7)} \quad P(CS) = P_r\left(Z_{ij} \le \frac{\delta^*}{\sqrt{\frac{\sigma^2_{(i)}}{n_{(i)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}}},\ i=1,\ldots,k-t;\ j=k-t+1,\ldots,k\right)$$

where

$$\text{(1.4.8)} \qquad Z_{ij} = \frac{\bar{X}_{(i)} - \bar{X}_{(j)}}{\sqrt{\frac{\sigma^2_{(i)}}{n_{(i)}} + \frac{\sigma^2_{(j)}}{n_{(j)}}}}$$

are components of the t(k-t)-variate normal $\underline{Z}$ with zero mean vector, unit variances, and correlation matrix,

$$\begin{pmatrix} A_{(k-t+1)} & B_{k-t+1,k-t+2} & \cdots & B_{k-t+1,k} \\ B_{k-t+2,k-t+1} & A_{(k-t+2)} & \cdots & B_{k-t+2,k} \\ \vdots & & & \\ B_{k,k-t+1} & B_{k,k-2} & \cdots & A_{(k)} \end{pmatrix}$$

where $A_{(\ell)} = (a_{ij}^{(\ell)})$ and $B_{r,s} = (b_{ij}^{(r,s)})$ are (k-t)x(k-t) matrices such that $B_{k-t+i,k-t+j} = B'_{k-t+j,k-t+i}$, where prime stands for matrix transpose, and for $i \neq j$, $i,j \neq \ell$, $1 \le \ell \le k$,

$$a_{ij}^{(\ell)} = \begin{cases} \{(1 + \frac{n_{(\ell)}}{n_{(i)}} \frac{\sigma^2_{(i)}}{\sigma^2_{(\ell)}})(1 + \frac{n_{(\ell)}}{n_{(j)}} \cdot \frac{\sigma^2_{(j)}}{\sigma^2_{(\ell)}})\}^{-\frac{1}{2}} \text{ if } 1 \leq i < \ell;\ 1 \leq j < \ell, \\ \{(1 + \frac{n_{(\ell)}}{n_{(i+1)}} \cdot \frac{\sigma^2_{(i+1)}}{\sigma_{(\ell)}})(1 + \frac{n_{(\ell)}}{n_{(j)}} \cdot \frac{\sigma^2_{(j)}}{\sigma^2_{(\ell)}})\}^{-\frac{1}{2}} \text{ if } \ell < i \leq k-1; \\ \qquad\qquad\qquad\qquad\qquad\qquad 1 \leq j < \ell, \\ \{(1 + \frac{n_{(\ell)}}{n_{(i+1)}} \cdot \frac{\sigma^2_{(i+1)}}{\sigma_{(\ell)}})(1 + \frac{n_{(\ell)}}{n_{(j+1)}} \cdot \frac{\sigma^2_{(j+1)}}{\sigma_{(\ell)}})\}^{-\frac{1}{2}} \quad \text{if} \\ \qquad\qquad\qquad\qquad \ell < i \leq k-1;\ \ell < j \leq k-1, \end{cases}$$

and

$$a_{ij}^{(\ell)} = a_{ji}^{(\ell)} \qquad \text{for all } i,j \neq \ell$$

$$a_{ii}^{(\ell)} = 1 \qquad i = 1,\ldots,k-t$$

and

$$b_{uv}^{(r,s)} = \delta_{uv}\{(1 + \frac{n_{(u)}}{n_{(r)}} \cdot \frac{\sigma^2_{(r)}}{\sigma^2_{(u)}})(1 + \frac{n_{(v)}}{n_{(s)}} \cdot \frac{\sigma^2_{(s)}}{\sigma^2_{(v)}})\}^{-\frac{1}{2}}$$

where δ_{uv} stands for the Kronecker delta function. By Lemma 1.2.1, and the choice of c, it follows that

$$P(CS) \geq P_r(Y_{ij} \geq c\{(n_0-1)(\frac{1}{Y_i} + \frac{1}{Y_j})\}^{-\frac{1}{2}};\ i=1,\ldots,k-t,\ j=k-t+1,\ldots,k)$$

where Y_{ij} are iid standard normal; Y_i and Y_j are iid chi-square random variate with (n_0-1) degree of freedom, and also $\{Y_{ij}\}$ and $\{Y_i, Y_j\}$ are stochastically independent. Hence

$$P(CS) \geq \int_0^\infty \ldots \int_0^\infty \prod_{i=1}^{t} \{\Phi(\frac{c}{\sqrt{(n_0-1)(\frac{1}{x} + \frac{1}{y_i})}})dG(x)\}^{k-t} dG(y_i)$$

This completes the proof.

It should be pointed out that when $t = 1$, it reduces to the result obtained by Rinott [80].

1.5 K-SAMPLE BEHRENS-FISHER PROBLEM

The Behrens-Fisher problem in its original simple version can be formulated as follows: Given two samples $X_{11},\ldots,X_{1n_1}$ and $X_{21},\ldots,X_{2n_2}$. It is assumed that the values of the first sample are generated from a normal distribution with mean μ_1 and variance σ_1^2 and that the values of the second samples come from a normal distribution with mean μ_2 and variance σ_2^2. The true values of μ's and σ's are not known and the sample sizes and the variances are possibly not equal. The problem consists in making inference about the actual value of the difference $\mu_1 - \mu_2$ of the means. So far, no entirely satisfactory test for the Behrens-Fisher problem has been derived. When $k = 2$, several solutions to this problem were provided (see Pfanzagl [73]). Unfortunately none of these methods is applicable for the case when $k \geq 3$.

In this section, we demonstrate that the procedure given in Section 1.2 provides a solution for the Behren-Fisher problem when $k \geq 3$.

Now let $\pi_1,\ldots,\pi_k$ be k independent normal populations with means $\mu_1,\ldots,\mu_k$ and variances $\sigma_1^2,\ldots,\sigma_k^2$, respectively. Suppose we are allowed to take $n_i (\geq 2)$ observations from each normal population π_i, $i = 1,\ldots,k$. Based on these observations, we wish to know whether μ_i are significantly different or not. The problem is to test the homogeneity of the means of the k normal populations. Let

n_0 be a fixed integer such that $2 \leq n_0 \leq \min_{1 \leq i \leq k} \{n_i\}$. Use the two-sample procedure given in Section 1.2. to obtain sample variances

$$S_i^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{ij} - \tilde{X}_i)^2, \quad i = 1,\ldots,k$$

where

$$\tilde{X} = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{ij},$$

and

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \qquad i = 1,\ldots,k.$$

Then our test rejects the hypothesis

$$H_0: \ \mu_1 = \ldots = \mu_k$$

when $\max_{1 \leq i \leq k} \bar{X}_i - \min_{1 \leq j \leq k} \bar{X}_j \geq cV$, where $V = \max_{1 \leq i \leq k} \frac{S_i^2}{n_i}$ and c is a constant such that the hypothesis of homogeneity will be rejected at the level α if the observed value of $\max_{1 \leq i \leq k} \bar{X}_i - \min_{1 \leq j \leq k} \bar{X}_j$ is greater than cV. Now using Theorem 1.2.3, we have the following theorem.

Theorem 1.5.1. For any α, $0 < \alpha < 1$, let $P^* = 1 - \frac{\alpha}{k}$ and let c be the constant determined by (1.2.29), then

$$\text{(1.5.1)} \qquad \max_{n_1,\ldots,n_k} \sup_{\Omega_0} P(H_0 \text{ is rejected}) < \alpha$$

where $\Omega_0 = \{\underline{\mu} = (\mu,\ldots,\mu), -\infty < \mu < \infty\}$.

<u>Proof</u>.

$$\sup_{\Omega_0} P(H_0 \text{ is rejected})$$

$$= \sup_{\Omega_0} P\left(\max_{1 \leq i \leq k} \bar{X}_i - \min_{1 \leq j \leq k} \bar{X}_j > cd \right)$$

$$= \sup_{\Omega_0} P\left(\bar{X}_j < \max_{1 \leq i \leq k} \bar{X}_i - cd, \quad \text{for some } j,\ 1 \leq j \leq k\right)$$

$$\leq k \sup_{\Omega_0} P\left(\bar{X}_k < \max_{1 \leq i \leq k} \bar{X}_i - cd\right)$$

$$= k\left(1 - \inf_{\Omega_0} P\left(\bar{X}_k \geq \max_{1 \leq i \leq k} \bar{X}_i - cd\right)\right)$$

$$\leq k \left(1 - \left(1 - \frac{\alpha}{k}\right)\right)$$

$$= \alpha.$$

CHAPTER 2

ON SOME NONPARAMETRIC SUBSET SELECTION PROCEDURES FOR THE LOCATION AND SCALE PARAMETERS

2.1 INTRODUCTION

In the past twenty years many papers have appeared on ranking and selection problems. As can be expected, most of this research has been devoted to rules which assume a specific distributional form of the underlying observations; e.g. normal, binomial, multinomial, etc. Barlow and Gupta [6] have considered the problem of selecting a subset containing the largest (smallest) quantile of a given order and a subset containing the largest (smallest) mean. They assume the observations from each population have a distribution which belongs to certain restricted family, e.g. IFR (Increasing Failure Rate) distributions, IFRA (Increasing Failure Rate Average) distributions, etc. Distribution-free selection procedures which are based on joint ranks of the observations, have been studied by Lehmann [62], Rizvi and Sobel [81], Bartlett and Govindarajulu [9], Puri and Puri [74] and Gupta and McDonald [43]. In [37], Gupta and Huang proposed some selection procedures based on one-sample Hodges-Lehmann estimates of the parameters for a class of distribution functions $F(x-\theta)$ where F and θ are both unknown. In this chapter, we discuss some subset selection procedures based on generalized U-statistics for selecting the location parameters. It should be pointed out that some related non-parametric selection procedures have also been discussed by Bhapkar and Gore [15].

In Section 2.3, we propose a subset selection rule for selecting the largest location parameter. It is assumed that all the observations are available i.e. it is the complete sample case. An upper bound for the supremum of the expected size of the selected subset is also given. In Section 2.4, we consider the case of selecting a subset which contains the population with the smallest scale parameter. In Section 2.5, we investigate the situation when the samples are trimmed.

2.2 NOTATIONS AND REQUIREMENTS

Let $\pi_1,\ldots,\pi_k$ be k independent populations with continuous cumulative functions $F_{\theta_1}(x),\ldots,F_{\theta_k}(x)$ respectively. The ordered θ's are denoted by $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$. It is assumed that there is no a priori information available about the correct pairing of the ordered $\theta_{[i]}$ and the k populations from which observations are taken. Any population whose parameter value equals $\theta_{[k]}$ ($\theta_{[1]}$) will be defined as a best population. The problem of primary interest is to define a procedure R which selects a subset of the k given populations that is small, never empty and large enough so that it contains the best population with probability at least P* ($\frac{1}{k} < P^* < 1$) regardless of the true configuration, i.e.

$$\inf_{\Omega} P(CS|R) \geq P^* \tag{2.2.1}$$

where Ω denote the space of all possible configurations.

Suppose we have k independent random samples, $X_{i1}, X_{i2}, \ldots, X_{in_i}$ say, of size n_i from population π_i with a continuous distribution F_i, i = 1,2,...,k. We form k-tuples by taking one observation from each sample. Put for i = 1,2,...,k,

$$U_i = \frac{1}{\prod_{i=1}^{k} n_i} \sum_{\alpha_1=1}^{n_1} \cdots \sum_{\alpha_k=1}^{n_k} \phi_i(X_{1\alpha_1}, \ldots, X_{k\alpha_k}) \tag{2.2.2}$$

with

$$\phi_i(x_1,\ldots,x_k) = \begin{cases} \frac{(j-1)_r}{(k-1)_r} - \frac{(k-j)_s}{(k-1)_s} & \text{if } x_i \text{ is the } j\underline{th} \text{ smallest among } x_1,\ldots,x_k \\ 0 & \text{otherwise,} \end{cases} \tag{2.2.3}$$

where $(h)_m = h(h-1)\ldots(h-m+1)$, $(h)_0 = 1$, r and s are two fixed integers such that $0 \leq r, s \leq k-1$ except for (r,s) = (0,0). Then U_i is called a generalized U-statistic. It should be pointed out that in case r = 0 and s = k-1, (2.2.3) reduces to

$$\phi_i(x_1,\ldots,x_k) = \begin{cases} 0 & \text{if } x_i < x_j \text{ for all } j \neq i \\ 1 & \text{otherwise.} \end{cases} \tag{2.2.4}$$

This leads to Bhapkar's V-statistics (see [15]) which has been offered for testing the hypothesis

(2.2.5) $$H_0: \; F_1 = \ldots = F_k$$

against the location alternatives

(2.2.6) H_ℓ: $F_i(x) = F(x-\theta_i)$ not all θ's being equal

or the scale alternatives

(2.2.7) H_s: $F_i(x) = F(\frac{x}{\sigma_i})$ not all σ's being equal for skew distribution F.

If r = s = k-1, then (2.2.3) reduces to

(2.2.8) $$\phi_i(x_1,\ldots,x_k) = \begin{cases} 1 & \text{if } x_i > x_j \text{ for all } j \neq i \\ -1 & \text{if } x_i < x_j \text{ for all } j \neq i \\ 0 & \text{otherwise.} \end{cases}$$

This leads to Deshpande's L-statistics which has been proposed for testing the hypothesis H_0 against H_ℓ for symmetric F.

Now we use the U-statistics for subset selection problem.

2.3 SELECTING A SUBSET CONTAINING THE POPULATION WITH THE LARGEST LOCATION PARAMETER

Suppose we have k independent populations $\pi_1,\pi_2,\ldots,\pi_k$ with continuous cumulative distribution functions $F(x-\theta_1)$, $F(x-\theta_2)$, ..., $F(x-\theta_k)$ respectively. Let $X_{i1},\ldots,X_{in_i}$ be n_i independent observations from population π_i, and let $n_i = r_iN$, $i = 1,2,\ldots,k$. Suppose we are interested in choosing a small but non-empty subset according to the procedure (to be defined below) such that the probability is at least P* that the selected subset contains the population with the largest value of θ. We define a rule based on U-statistics as follows:

R_5: Select population π_i if and only if

$$(2.3.1) \qquad U_i \geq \max_{1 \leq j \leq k} U_j - \frac{c_5}{\sqrt{N}} \frac{k}{k-1} \sqrt{\frac{\eta(r,s)}{n_i}}$$

where

$$(2.3.2) \qquad \eta(r,s) = \frac{r^2}{(2r+1)(r+1)^2} + \frac{s^2}{(2s+1)(s+1)^2} + \frac{2}{(r+1)(s+1)} - \frac{2r!s!}{(r+s+1)!}$$

and $c_1 = c_1(k,P^*,r,s,n_1,\ldots,n_k)$ is the smallest non-negative number such that the basic probability requirement (2.2.1) is satisfied.

For the evaluation of the infimum of $P(CS|R_5)$, we need the following lemma which is due to Sugiura [94] and is stated below without proof.

Lemma 2.3.1. Assume the sequence of distribution

$$(2.3.3) \qquad F_i^{(N)}(x) = F(x-N^{-\frac{1}{2}}\theta_i)$$

of independent random variables X_{ij}, $j = 1,\ldots,n_i$ for each N, where $n_i = r_i N$ with r_i being a positive constant independent of N, $i = 1,2,\ldots,k$. Suppose that the distribution F(x) possesses a continuous derivative f(x) except for a set of measure zero and further there exists a function g(x) such that

$$(2.3.4) \qquad \int_{-\infty}^{\infty} g(x)dF(x) < \infty$$

and that

$$(2.3.5) \qquad \left|\frac{f(x+h)-f(x)}{h}\right| \leq g(x)$$

holds for every x and any sufficiently small h. Let $\underset{\sim}{U}' = (U_1,\ldots,U_k)$, $\underset{\sim}{1}' = (1,\ldots,1)$. Then as $N \to \infty$, the limiting distribution of $\sqrt{N}\,(\underset{\sim}{U} - (\frac{1}{r+1} - \frac{1}{s+1})\underset{\sim}{1})$ is k-variate normal with mean vector $\underset{\sim}{\mu}' = (\mu_1,\ldots,\mu_k)$ and covariance matrix (σ_{ij}), where

$$(2.3.6) \qquad \mu_i = \frac{\lambda_k(r,s)}{k-1} \sum_{\alpha=1}^{k} (\theta_i - \theta_\alpha), \quad i = 1,2,\ldots,k,$$

with

$$(2.3.7) \qquad \lambda_k(r,s) = \int_{-\infty}^{\infty} \{rF^{r-1}(x) + s(1-F(x))^{s-1}\} f(x)\,dF(x)$$

and

$$(2.3.8) \qquad \sigma_{ij} = \frac{\eta(r,s)}{(k-1)^2} \left\{ \sum_{\alpha=1}^{k} \frac{1}{r_\alpha} - k\left(\frac{1}{r_i} + \frac{1}{r_j}\right) + \frac{k^2 \delta_{ij}}{r_i} \right\}.$$

We now prove the following theorem regarding the infimum of $P(CS|R_5)$ when the sample sizes are large.

Theorem 2.3.1. Suppose that the assumptions given in Lemma 2.3.1 hold. Then for large N,

$$(2.3.9) \qquad \min_{n_1,\ldots,n_k} \inf_{\Omega} P(CS|R_5) \sim \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \Phi\left(\frac{\alpha_j}{\sqrt{1-\alpha_j^2}}(x+c_5)\right) d\Phi(x)$$

where $\alpha_j = (1 + \frac{n_{[k]}}{n_{[j]}})^{-\frac{1}{2}}$, $j = 1,\ldots,k-1$.

<u>Proof</u>. Let $\underline{\theta} = (\theta_1,..,\theta_k) \in \Omega$ and let $n_{(i)} = r_{(i)}N$ be the unknown sample size associated with the population $\pi_{(i)}$ whose parameter is $\theta_{[i]}$. Then

$$P_{\underline{\theta}}(CS|R_5) = P_{\underline{\theta}}(U_{(k)} \geq U_{(j)} - \frac{c_5}{\sqrt{N}}\frac{k}{k-1}\sqrt{\frac{n(r,s)}{n_{(k)}}},\quad j = 1,\ldots,k-1)$$

$$(2.3.10)\qquad = P_{\underline{\theta}}(Y_j \leq \frac{\lambda_k(r,s)(\theta_{[k]}-\theta_{[j]})}{\{n(r,s)(\frac{1}{n_{(k)}} + \frac{1}{n_{(j)}})\}^{\frac{1}{2}}} + \frac{c_5}{(1 + \frac{n_{(k)}}{n_{(j)}})^{\frac{1}{2}}},\quad j = 1,\ldots,k-1)$$

where the (k-1)-dimensional random vector with components

$$(2.3.11)\qquad Y_j = \frac{\sqrt{N}(U_{(j)}-U_{(k)}) + k(k-1)^{-1}\lambda_k(r,s)(\theta_{[k]}-\theta_{[j]})}{k(k-1)^{-1}\{n(r,s)(\frac{1}{r_{(k)}} + \frac{1}{r_{(j)}})\}^{\frac{1}{2}}}$$

is, as $N \to \infty$, distributed as a (k-1)-variate normal, $(Z_1,\ldots,Z_{k-1})$, say, with zero mean vector, unit variances and correlation matrix, where

$$(2.3.12)\qquad \rho_{ij}^{(k)} = \{(1 + \frac{n_{(k)}}{n_{(i)}})(1 + \frac{n_{(k)}}{n_{(j)}})\}^{-\frac{1}{2}},\quad i,j = 1,\ldots,k-1,\ i \neq j.$$

For any given association between $\{n_1,\ldots,n_k\}$ and $\{n_{(1)},\ldots,n_{(k)}\}$, we see from (2.3.10) that the infimum of $P(CS|R_5)$ is attained when $\theta_{[1]} =\ldots= \theta_{[k]}$. Thus as N large, the infimum we seek in (2.3.9) is given by

$$(2.3.13)\qquad \min_{n_{(1)},\ldots,n_{(k)}} P_r\left(Z_j \le c_5\left(1 + \frac{n_{(k)}}{n_{(j)}}\right)^{-\frac{1}{2}},\quad j = 1,\ldots,k-1\right).$$

It follows that (2.3.13) is minimized when $n_{(i)} = n_{[i]}$, $i = 1,\ldots,k$. By using the transformation (1.2.14), (2.3.13) can now be expressed as

$$\int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \Phi\left(\frac{\alpha_j}{\sqrt{1-\alpha_j^2}}\,(c_5 - x)\right) d\Phi(x).$$

Replace x by -x' and drop the prime, gives the right hand member of (2.3.9). This completes the proof of the theorem.

<u>Remark 2.3.1</u>. It should be pointed out that when $n_1 = \ldots = n_k$, the right hand member of (2.3.9) reduces to

$$(2.3.14)\qquad \int_{-\infty}^{\infty} \Phi^{k-1}(x+c_5)\,d\Phi(x).$$

Let S denote the size of the selected subset. The expected value of S when R_5 is used is given by

$$(2.3.15)\qquad E(S|R_5) = \sum_{i=1}^{k} P(\pi_{(i)} \text{ is selected } |R_5)$$

$$= \sum_{i=1}^{k} P_r\left(U_{(i)} \ge \max_{1\le j\le k} U_{(j)} - \frac{c_5}{\sqrt{N}}\,\frac{k}{k-1}\sqrt{\frac{n(r,s)}{n_{(i)}}}\right)$$

$$= \sum_{i=1}^{k} P_r\left(\max_{\substack{1\le j\le k\\ j\ne i}} (U_{(j)} - U_{(i)}) \le \frac{c_5}{\sqrt{N}}\;\frac{k}{k-1}\;\sqrt{\frac{n(r,s)}{n_{(i)}}}\right)$$

Theorem 2.3.2. Suppose that the assumptions given in Lemma 2.3.1 hold, then for large N,

$$(2.3.16)\qquad \max_{n_1,\ldots,n_k} \sup_{\Omega} E(S|R_5) \le \frac{1}{k-1} \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \Phi\left(\sqrt{\frac{n_{[k]}}{n_{[i]}+n_{[j]}}}\, c_5\right).$$

Proof: Since

$$(2.3.17)\qquad P_r\Big(\max_{\substack{1\le j\le k \\ j\neq i}} (U_{(j)}-U_{(i)}) \le \frac{c_5}{\sqrt{N}} \frac{k}{k-1} \sqrt{\frac{\eta(r,s)}{n_{(i)}}}\Big) \le P_r\Big(U_{(j)}-U_{(i)} \le \frac{c_5}{\sqrt{N}} \frac{k}{k-1} \sqrt{\frac{\eta(r,s)}{n_{(i)}}}\Big)$$

for all $j \neq i$, it follows that

$$(2.3.17)\qquad P_r\Big(\max_{\substack{1\le j\le k \\ j\neq i}} (U_{(j)}-U_{(i)}) \le \frac{c_5}{\sqrt{N}} \frac{k}{k-1} \sqrt{\frac{\eta(r,s)}{n_{(i)}}}\Big) \le \frac{1}{k-1} \sum_{\substack{j=1 \\ j\neq i}}^{k} P_r\Big(U_{(j)}-U_{(i)} \le \frac{c_5}{\sqrt{N}} \frac{k}{k-1} \sqrt{\frac{\eta(r,s)}{n_{(i)}}}\Big).$$

For large N, the right hand member of (2.3.17) approaches

$$\frac{1}{k-1} \sum_{\substack{j=1 \\ j\neq i}}^{k} \Phi\Big(c_5\Big(1 + \frac{n_{(i)}}{n_{(j)}}\Big)^{-\frac{1}{2}} + \eta(r,s)^{-\frac{1}{2}}\Big(\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}}\Big)^{-\frac{1}{2}} \lambda_k(r,s)(\theta_{[i]}-\theta_{[j]})\Big).$$

Therefore, for large N,

$$(2.3.18)\quad E(S|R_5) \le \frac{1}{k-1} \sum_{i=1}^{k} \sum_{\substack{j=1\\ j\ne i}}^{k} \Phi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}}) + \eta(r,s)^{-\frac{1}{2}}(\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}})^{-\frac{1}{2}} \lambda_k(r,s)(\theta_{[i]} - \theta_{[j]}))$$

$$= \frac{1}{k-1} Q_1 \quad \text{(say)}.$$

Under the configuration (1.2.19), Q_1 can be expressed as follows:

$$(2.3.19)\quad Q_1 = \sum_{i=1}^{m} \{ \sum_{\substack{j=1\\ j\ne i}}^{m} \Phi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}})^{-\frac{1}{2}}) + \sum_{j=m+1}^{k} \Phi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}})^{\frac{1}{2}} + $$

$$+ \eta(r,s)^{-\frac{1}{2}}(\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}})^{-\frac{1}{2}} \lambda_k(r,s)(\theta-\theta_{[j]}))\}$$

$$+ \sum_{i=m+1}^{k} \{ \sum_{j=1}^{m} \Phi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}})^{\frac{1}{2}} + \eta(r,s)^{-\frac{1}{2}}(\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}})^{-\frac{1}{2}} \lambda_k(r,s)(\theta_{[i]} - \theta))$$

$$+ \sum_{j=m+1}^{k} \Phi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}})^{\frac{1}{2}} + \eta(r,s)^{-\frac{1}{2}}(\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}})^{-\frac{1}{2}} \lambda_k(r,s)(\theta_{[i]} - \theta_{[j]}))\}$$

Differentiate Q_1 with respect to θ, we get

$$(2.3.20)\quad \frac{\partial Q_1}{\partial \theta} = \sum_{i=1}^{m} \sum_{j=m+1}^{k} \eta(r,s)^{-\frac{1}{2}} \lambda_k(r,s) \{ \varphi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}})^{\frac{1}{2}} - \eta(r,s)^{-\frac{1}{2}} (\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}})^{-\frac{1}{2}} \lambda_k(r,s)(\theta_{[j]}-\theta))$$

$$- \varphi(c_5(\frac{n_{[k]}}{n_{(i)}+n_{(j)}})^{\frac{1}{2}} + \eta(r,s)^{-\frac{1}{2}}(\frac{1}{n_{(i)}} + \frac{1}{n_{(j)}})^{-\frac{1}{2}} \lambda_k(r,s)(\theta_{[j]}-\theta))\}.$$

Since if a is positive then $\varphi(a-x) \geq \varphi(a+x)$ for all nonnegative x. It follows from (2.3.21) that Q_1 is nondecreasing in θ. Thus inequality (2.3.16) follows.

2.4. SELECTING A SUBSET CONTAINING THE POPULATION WITH THE SMALLEST SCALE PARAMETER

Let $\pi_1,\ldots,\pi_k$ denote k independent populations with continuous cumulative distribution $F(\frac{x}{\sigma_1}),\ldots,F(\frac{x}{\sigma_k})$ respectively. The functional form of F and the parameters are assumed to be unknown. Let $X_{i1},X_{i2},\ldots,X_{in_i}$ be n_i random samplestaken from population π_i, $i = 1,\ldots,k$. Corresponding to the statistic (2.2.2), we shall put at this time for $i = 1,2,\ldots,k$

$$(2.4.1)\qquad U_i = \frac{1}{\prod_{i=1}^{k} n_i} \sum_{\alpha_1=1}^{n_1} \cdots \sum_{\alpha_k=1}^{n_k} \phi_i(X_{1\alpha_1},\ldots,X_{k\alpha_k})$$

with

$$(2.4.2)\qquad \phi_i(X_1,\ldots,X_k) = \begin{cases} \dfrac{(j-1)_r}{(k-1)_r} + \dfrac{(k-j)_s}{(k-1)_s} & \text{if } X_i \text{ is the } j\text{th smallest among } X_1,\ldots,X_k, \\ 0 & \text{otherwise} \end{cases}$$

where $0 \leq r,s \leq k-1$ except for $(r,s) = (0,0)$ and $(1,1)$. In case $r = s = k-1$, (2.4.2) reduces to

$$(2.4.3)\qquad \phi_i(X_1,\ldots,X_k) = \begin{cases} 1 & \text{if } X_i < X_j \text{ or } X_j > X_i \text{ for any } j \neq i \\ 0 & \text{otherwise} \end{cases}$$

which leads to Deshpande's D-statistics which has been proposed for testing the hypothesis H_0 against H_s. If $r = 0$, $s = k-1$, $\phi_i(X_1,\ldots,X_k)$ is given by (2.2.4) which leads to Bhapkar's V-statistics. Now we define a subset selection rule based on these U-statistics as follows:

R_6: Select population π_i if and only if

$$U_i \le \min_{1\le j\le k} U_j + \frac{c_6}{\sqrt{N}} \cdot \frac{k}{k-1} \sqrt{\frac{L(r,s)}{n_i}} \tag{2.4.4}$$

where

$$L(r,s) = \frac{r^2}{(2r+1)(r+1)^2} + \frac{s^2}{(2s+1)(s+1)^2} - \frac{2}{(r+1)(s+1)} + \frac{2r!s!}{(r+s+1)!}, \tag{2.4.5}$$

and c_6 is the smallest nonnegative constant so that (2.2.1) is satisfied.

The following result is due to Sugiura [94].

<u>Lemma 2.4.1.</u> Assume the sequence of distributions

$$F_i(x) = F\left(\frac{x}{a+N^{-\frac{1}{2}}\sigma_i}\right) \tag{2.4.6}$$

of independent random variables X_{ij}, $j = 1,\ldots,n_i$ for each N where $n_i = r_iN$ with r_i being a positive constant independent of N, $i = 1,\ldots,k$. Suppose that the distribution $F(x)$ has a derivative $f(x)$ except for a set of measure zero and further that there exists a function $g(x)$ such that

(2.4.7) $$\int_{-\infty}^{\infty} xg(x)dF(x) < \infty$$

and that

(2.4.8) $$\left|\frac{f(x+h)-f(x)}{h}\right| \leq g(x)$$

holds for every x and any sufficiently small h. Let $\underline{U}'$ and $\underline{1}'$ be defined as in Lemma 2.3.1. Then as $N \to \infty$, the limiting distribution of $\sqrt{N}(\underline{U} - (\frac{1}{r+1} + \frac{1}{s+1})\underline{1})$ is k-variate normal with mean vector $\underline{\mu}' = (\mu_1,\ldots,\mu_k)$ and covariance matrix (ρ_{ij}) defined as follows:

(2.4.9) $$\mu_i = \frac{\xi(r,s)\sum_{\alpha=1}^{k}(\sigma_i-\sigma_\alpha)}{a(k-1)}, \quad i = 1,\ldots,k$$

with

(2.4.10) $$\xi(r,s) = \int_{-\infty}^{\infty} xf(x)\{rF^{r-1}(x)-s(1-F(x))^{s-1}\}dF(x),$$

and

(2.4.11) $$\sigma_{ij} = \frac{L(r,s)}{(k-1)^2}\{\sum_{\alpha=1}^{k}\frac{1}{r_\alpha} - k(\frac{1}{r_i} + \frac{1}{r_j}) + \frac{k^2\delta_{ij}}{r_i}\}.$$

Now

$$P(CS|R_6) = P_r(U_{(1)} \leq \min_{2\leq i\leq k} U_{(i)} + \frac{c_6}{\sqrt{N}}\frac{k}{k-1}\sqrt{\frac{L(r,s)}{n_{(1)}}})$$

$$= P_r(Y_i \leq \frac{\xi(r,s)(\sigma_{[i]}-\sigma_{[1]})}{\sqrt{L(r,s)(\frac{1}{n_{(1)}} + \frac{1}{n_{(i)}})}} + \frac{c_6}{\sqrt{1 + \frac{n_{(1)}}{n_{(i)}}}};\ i=2,\ldots,k)$$

where the limiting distribution of $\underline{Y}$ with component

$$Y_i = \frac{\sqrt{N}(U_{(1)}-U_{(i)}) + \frac{\xi(r,s)}{a(k-1)} k(\sigma_{[i]}-\sigma_{[1]})}{(\frac{k}{k-1}) \sqrt{L(r,s)(\frac{1}{r_{(1)}} + \frac{1}{r_{(i)}})}}$$

is (k-1)-variate normal with zero mean vector, unit variance and correlation matrix given by (2.3.12). Using the analogous arguments as in the proof of Theorem 2.3.1, we obtain the following theorem.

Theorem 2.4.1. Suppose that the assumptions given in Lemma 2.4.1 hold. Then for large N,

$$\min_{n_1,\ldots,n_k} \inf_{\Omega_2} P(CS|R_6) \sim \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \Phi\left(\frac{\alpha_j}{\sqrt{1-\alpha_j^2}}(x+c_6)\right) d\Phi(x)$$

where α_j is defined in Theorem 2.3.1 and Ω_2 denote the set of all possible configurations of $(\sigma_1,\ldots,\sigma_k)$.

Similarly, one can show that

Theorem 2.4.2. For large N,

$$\max_{n_1,\ldots,n_k} \sup_{\Omega_2} E(S|R_6) \leq \frac{1}{k-1} \sum_{i=1}^{k} \sum_{j\neq i} \Phi\left(\sqrt{\frac{n_{[k]}}{n_{[i]}+n_{[j]}}}\, c_2\right)$$

2.5. SELECTING A SUBSET CONTAINING THE POPULATION WITH THE LARGEST LOCATION PARAMETER: TRIMMED SAMPLES CASE

Consider a single sample of size n from a distribution with absolutely continuous distribution function $F(x-\theta)$. For distribution with heavy tails, Tukey [96] proposed and investigated the α-trimmed mean as an estimate of θ which is based on the middle $n-2[n\alpha]$

observations. Later Huber [54] considered essentially the class of maximum likelihood estimates and found that the trimmed mean minimizes the maximum variance over various classes of contaminated distributions. In [16], Bickel studied the asymptotic relative efficiency properties of the α-trimmed mean relative to the mean for the class of continuous distributions with symmetric unimodal densities. Recently, Hettmansperger [51] has shown that one can increase the asymptotic relative efficiency in Pitman's sense of the Mann-Whitney test for some distributions with heavy tails by using the trimmed samples instead of the complete sample. Here a subset selection procedure based on trimmed samples is proposed and studied for distributions with heavy tails.

Let $X_{i[1]} \leq \cdots \leq X_{i[n_i]}$ be the ordered statistics from absolutely continuous distribution $F_i(x) = F(x-\theta_i)$, $i = 1,\ldots,k$, where $F(x)$ has a symmetric density of unknown functional form. We further assume that for $0 < \alpha < \frac{1}{2}$ that $F(x)$ is continuously differentiable in some neighborhood of the unique quantile q_α and $q_{1-\alpha}$ of order α and $1-\alpha$ respectively. In this section, we propose a subset selection rule based only on the middle $n_i-2\lambda_i$ ordered observations $X_{i[\lambda_i+1]} \leq \cdots \leq X_{i[n-\lambda_i]}$, from population π_i, $i = 1,\ldots,k$, where $\lambda_i = [n_i\lambda]$ denote the largest integer not exceeding $n_i\lambda$. We assume further that $n_i = r_iN$, $i = 1,\ldots,k$. Define a U-statistic as follows:

$$(2.5.1)\quad U_i^{(\alpha)} = \frac{1}{\prod_{j=1}^{k}(n_j-2\lambda_j)} \sum_{\alpha_1=\lambda_1+1}^{n_1-\lambda_1} \cdots \sum_{\alpha_k=\lambda_k+1}^{n_k-\lambda_k} \phi_i(X_{1\alpha_1},\ldots,X_{k\alpha_k}),$$

$$i = 1,\ldots,k$$

where ϕ_i is defined by (2.2.4).

Now we define a selection procedure as follows:

R_6': Retain the population π_i in the selected subset if and only if

$$(2.5.2) \qquad U_i^{(\alpha)} \geq \max_{1\leq j\leq k} U_j^{(\alpha)} - \frac{c_6'}{\sqrt{(2k-1)r_k(N-2\lambda)}}$$

where $n_i = r_i N$, $i = 1,\ldots,k$, $\lambda = \sum_{i=1}^{k} \lambda_i$.

Suppose that the assumptions stated in Lemma 2.3.1 are satisfied. The random vector $\sqrt{N-2\lambda}\ (\underline{U}^{(\alpha)} - \frac{1}{k}\underline{1})$ has the joint asymptotic normal distribution $N(\underline{\mu},\Sigma_\alpha)$, where

$$\mu_i = (1-2\alpha)^{\frac{1}{2}-k}\int_{q_\alpha}^{q_{1-\alpha}} \{F(x)-\alpha\}^{k-2} f(x)dF(x) \sum_{j=1}^{k} (\theta_i-\theta_j),$$

$i = 1,\ldots,k$, and $\Sigma_\alpha = (\omega_{ij})$,

$$\omega_{ij} = \frac{\beta}{k^2(2k-1)} \{\sum_{\beta=1}^{k} \frac{1}{r_\beta} + \frac{k^2\delta_{ij}}{r_i} - \frac{k}{r_i} - \frac{k}{r_j}\}$$

where

$$\beta = (k-1)^2(1-2\alpha) + (2k-1)(k^2-2k+2)\alpha(1-\alpha) + 2(k-1)(2k-1)\alpha^2.$$

It follows that

$$P(CS|R_6') = P_r(U_{(k)}^{(\alpha)} \geq \max_{1\leq j\leq k} U_{(j)}^{(\alpha)} - \frac{c_6'}{\sqrt{(2k-1)r_{(k)}(N-2\lambda)}})$$

$$= P_r(Y_j \leq \frac{c_6'}{\sqrt{1+\frac{n_{(k)}}{n_{(j)}}}} + \frac{k(\theta_{[k]}-\theta_{[j]})\gamma(\alpha)}{\frac{1}{2k-1}(\frac{1}{r_{(k)}}+\frac{1}{r_{(j)}})},\ j = 1,\ldots,k-1)$$

where

$$\gamma(\alpha) = (1-2\alpha)^{\frac{1}{2}-k} \int_{q_\alpha}^{1-q_\alpha} \{F(x)-\alpha\}^{k-2} f(x)\,dF(x),$$

and the limiting distribution of $(Y_1,\ldots,Y_{k-1})$ with component

$$Y_j = \frac{\sqrt{N-2\lambda}\,(U_{(j)}^{(\alpha)}-U_{(k)}^{(\alpha)})+k(\theta_{[k]}-\theta_{[j]})\gamma(\alpha)}{\sqrt{\frac{1}{2k-1}\left(\frac{1}{r_{(k)}}+\frac{1}{r_{(j)}}\right)}}, \quad j = 1,\ldots,k-1$$

is $(k-1)$-variate normal with zero mean vector, unit variances and correlation matrix given by (2.3.12). This implies that when N is large,

$$\min_{n_1,\ldots,n_k} \inf_{\Omega_1} P(CS|R_6) \sim \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \Phi\left(\frac{\alpha_j}{\sqrt{1-\alpha_j}}(x+c_6')\right) d\Phi(x).$$

Similarly, for large N, the inequality (2.3.16) holds.

CHAPTER III

ON SUBSET SELECTION PROCEDURES FOR POISSON PROCESSES AND SOME APPLICATIONS TO THE BINOMIAL AND MULTINOMIAL SELECTION PROBLEMS

3.1 INTRODUCTION

The Poisson process arises in many applications, especially as a model for arrivals at a store, for the arrivals of calls at a telephone exchange, for the arrivals of radioactive particles at a Geiger counter, etc. In this chapter, the problem of selecting a subset of k Poisson processes including the best which is associated with the smallest value of the mean arrival time is discussed. Some subset selection procedures are proposed and studied. An application of these procedures to the subset selection problem for the largest probability of a success of k binomial populations, whose parameters are unknown, is considered. Results are also applied to the problem of selecting the largest cell probability from a multinomial distribution, again the cell probabilities being unknown. It should be pointed out that single sample subset selection procedures have been considered by Gupta and Nagel [44], Gupta and Huang [39] and Goel [28]. Some parallel selection procedures have been discussed by Alam [1]. Recently Goel [29] also proposed a subset selection procedure for Poisson processes.

Let $\pi_1,\ldots,\pi_k$ be k Poisson processes with mean arrival times $\lambda_1^{-1},\ldots,\lambda_k^{-1}$, respectively. Let $\lambda_{[1]} \leq \ldots \leq \lambda_{[k]}$ denote the ordered set of the values $\lambda_1,\ldots,\lambda_k$. The process associated with $\lambda_{[k]}$ ($\lambda_{[1]}$) is defined to be the best process. Assume λ's are unknown and that there is no a priori information available about the correct pairing of the ordered $\lambda_{[i]}$ values and the k given processes. Our problem is to define a selection procedure which satisfies the P*-condition.

In Section 3.2, some subset selection rules for selecting a subset containing the process with largest value $\lambda_{[k]}$ are proposed. Three different kind of sampling rules are considered. The probability of a correct selection is evaluated. Some properties of the proposed selection rules are discussed. Section 3.3 deals with the analogous problem of selecting the process for which the associated value λ is the smallest. In Section 3.4, applications to binomial and multinomial selection problems are considered.

3.2 SELECTION PROCEDURES FOR PROCESS ASSOCIATED WITH $\lambda_{[k]}$

In this section, four different selection rules are proposed.

(A) PROCEDURE R_7 AND ITS PROPERTIES

Let $X_1(t),\ldots,X_k(t)$ denote the number of arrivals from processes $\pi_1,\ldots,\pi_k$ during time t, respectively. Let $X_{(i)}(t)$ and $\pi_{(i)}$ be associated with $\lambda_{[i]}$, $i = 1,\ldots,k$. Let N be a fixed positive integer. We propose a subset selection rule as follows:

R_7: Observe the processes until $\max_{1 \leq i \leq k} X_i(t) = N$. Select process π_i if and only if

(3.2.1) $$X_i(t) \geq N - c_7$$

where $c_7 = c_7(k, P^*, N)$ is the smallest non-negative integer for which the P^*-condition is satisfied.

Before we derive some properties of the selection rule, we introduce some definitions. Let Ω denote the set of all k-tuples $\underline{\lambda} = (\lambda_1, \ldots, \lambda_k)$, $\lambda_i > 0$, $i = 1, \ldots, k$. Define

(3.2.2) $$p_{\underline{\lambda}}(i) = P_{\underline{\lambda}}(\pi_{(i)} \text{ is selected } | R).$$

Definition 3.2.1. A rule R is said to be (reverse) strongly monotone in $\pi_{(i)}$ (see Santner [83]) if

$$p_{\underline{\lambda}}(i) \text{ is } \begin{cases} (\downarrow)\uparrow \text{ in } \lambda_{[i]} \text{ when all other components of } \underline{\lambda} \text{ are fixed.} \\ (\uparrow)\downarrow \text{ in } \lambda_{[j]}\ (j \neq i) \text{ when all other components of } \underline{\lambda} \text{ are fixed.} \end{cases}$$

Gupta [35] has proved that the subset selection rules which he studied possess the properties of monotonicity and unbiasedness. Recall these definitions:

Definition 3.2.2. The rule R is (reverse) monotone means for all $1 \leq i < j \leq k$, and $\underline{\lambda} \in \Omega$,

(3.2.3) $$p_{\underline{\lambda}}(i)(\geq) \leq p_{\underline{\lambda}}(j).$$

Definition 3.2.3. The rule R is unbiased means for all $1 \leq i < k$ and $\underline{\lambda} \in \Omega$,

$$P_{\underline{\lambda}}(\text{R does not select } \pi_{(i)}) \geq P_{\underline{\lambda}}(\text{R does not select } \pi_{(k)})$$

Remark 3.2.1. (1) If a rule R is (reverse) strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$, then R is monotone and

$$\inf_{\Omega} P(CS|R) = \inf_{\Omega_0} P(CS|R) \tag{3.2.4}$$

where $\Omega_0 = \{\underline{\lambda} = (\lambda,\ldots,\lambda)$, for all possible values of $\lambda\}$.

(2) If R is monotone, then it is unbiased.

Let $T_i(N)$ denote the waiting time for N arrivals for the process π_i, $i = 1,\ldots,k$. $T_i(N)$ is distributed according to gamma distribution with density given by

$$f_{i,N}(t) = \frac{\lambda_i^N}{\Gamma(N)} t^{N-1} e^{-\lambda_i t}, \quad t > 0. \tag{3.2.5}$$

Let $T_{(i)}(N)$ denote the unknown waiting time for N arrivals for the process $\pi_{(i)}$, $i = 1,\ldots,k$. Then for any $\underline{\lambda} \in \Omega$,

$$p_{\underline{\lambda}}(i) = P_{\underline{\lambda}}(X_{(i)}(t) \geq N-c_7)$$

$$= P_{\underline{\lambda}}(T_{(i)}(N-c_7) \leq \min_{1 \leq j \leq k} T_{(j)}(N))$$

$$= \int_0^\infty \prod_{\substack{j=1 \\ j \neq i}}^{k} \{1-G_N(\frac{\lambda_{[j]}}{\lambda_{[i]}} t)\} dG_{N-c_7}(t), \tag{3.2.6}$$

where

$$G_r(x) = \int_0^x \frac{1}{\Gamma(r)} t^{r-1} e^{-t} dt. \tag{3.2.7}$$

It follows from (3.2.6) that procedure R_7 is strongly monotone in $\pi_{(i)}$ for all i = 1,...,k. Furthermore

(3.2.8) $$\inf_{\Omega} P(CS|R_7) = \inf_{\Omega_0} P(CS|R_7) = \int_0^\infty \{1-G_N(t)\}^{k-1} dG_{N-c_7}(t)$$

which is independent of the common unknown parameter. Hence we have proved the following theorem.

Theorem 3.2.1. The procedure R_7 is strongly monotone in $\pi_{(i)}$ for all i = 1,...,k, and the infimum of the probability of a correct selection occurs when all the processes are identical and the infimum does not depend on the common unknown parameter.

Remark 3.2.2. In order to find the selection constant c_7 so as to satisfy the P*-condition, it suffices to choose c_7 of (3.2.1) to be the smallest integer greater than or equal to x where

(3.2.8) $$\int_0^\infty \{1-G_N(t)\}^{k-1} dG_{N-x}(t) = P^*.$$

Let S denote the size of the selected subset when procedure R_7 is used. S is a random variable which takes values 1,2,...,k. The expected value of S is given by

$$E_{\underline{\lambda}}(S|R_7) = \sum_{i=1}^{k} P_{\underline{\lambda}}(\pi_{(i)} \text{ is selected } |R_7)$$

(3.2.9) $$= \sum_{i=1}^{k} \int_0^\infty \prod_{\substack{j=1 \\ j\neq i}}^{k} \{1-G_N(\frac{\lambda_{[j]}}{\lambda_{[i]}} t)\} dG_{N-c_7}(t).$$

It will now be shown that the maximum of $E_{\underline{\lambda}}(S|R_7)$ takes place when all the parameters λ_i are equal. If we set the m smallest parameter $\lambda_{[i]}$ $(1 \leq m < k)$ equal to a common value λ (say), we obtain from (3.2.9) that

$$(3.2.10)\quad E_{\underline{\lambda}}(S|R_7) = m\int_0^\infty \{1-G_N(t)\}^{m-1} \cdot \prod_{j=m+1}^{k} \{1-G_N(\frac{\lambda_{[j]}}{\lambda} t)\} dG_{N-c_7}(t)$$

$$+ \sum_{i=m+1}^{k} \int_0^\infty \{1-G_N(\frac{\lambda}{\lambda_{[i]}} t)\} \prod_{\substack{j=m+1 \\ j\neq i}}^{k} \{1-G_N(\frac{\lambda_{[j]}}{\lambda_{[i]}} t)\} dG_{N-c_7}(t).$$

we now show that the right hand member of (3.2.10) is a nondecreasing function of λ for $k^{-1} < P^* < 1$. Since this holds for integer $m < k$, this proves that its maximum value occurs when $\lambda = \lambda_{[k]}$, and the desired result will follow. To show that $E_{\underline{\lambda}}(S|R_7)$ is monotone, we differentiate $E_{\underline{\lambda}}(S|R_7)$ with respect to λ and show that the derivative is positive for $k^{-1} < P^* < 1$. Differentiation gives

$$(3.2.11)\quad \frac{\partial}{\partial\lambda} E_{\underline{\lambda}}(S|R_7) = m \sum_{i=m+1}^{k} \int_0^\infty \{1-G_N(t)\}^{m-1} \prod_{\substack{j=m+1 \\ j\neq i}}^{k} \{1-G_N(\frac{\lambda_{[j]}}{\lambda} t)\}$$

$$f_{i,N}(\frac{\lambda_{[i]}}{\lambda} t)\frac{\lambda_{[i]}}{\lambda^2} t dG_{N-c_7}(t)$$

$$- m \sum_{i=m+1}^{k} \int_0^\infty \{1-G_N(\frac{\lambda}{\lambda_{[i]}} t)\}^{m-1} \prod_{\substack{j=m+1 \\ j\neq i}}^{k} \{1-G_N(\frac{\lambda_{[j]}}{\lambda_{[i]}} t)\}$$

$$f_{i,N}(\frac{\lambda}{\lambda_{[i]}} t)\frac{t}{\lambda_{[i]}} dG_{N-c_7}(t).$$

If we let $\lambda t = \lambda_{[i]} t'$ in the second integral and drop primes then (3.2.11) becomes

$$(3.2.12)\quad \frac{\partial}{\partial\lambda} E_{\underline{\lambda}}(S|R_7) = m \sum_{i=m+1}^{k} \{(1-G_N(t))^{m-1} \prod_{\substack{j=m+1 \\ j\neq i}}^{k} (1-G_N(\frac{\lambda_{[j]}}{\lambda} t))$$

$$\frac{1}{\Gamma(N)} \frac{1}{\Gamma(N-c_7)} (\frac{\lambda_{[i]}}{\lambda} t)^N \cdot$$

$$\frac{1}{\lambda} t^{N-c_7-1} e^{-t(1 + \frac{\lambda_{[i]}}{\lambda})} \}\{1-(\frac{\lambda}{\lambda_{[i]}})^{c_7+1}\} dt$$

$$\geq 0.$$

Hence we have proved the following theorem.

Theorem 3.2.2.

$$(3.2.13)\quad \sup_{\Omega} E_{\underline{\lambda}}(S|R_7) = k \int_0^\infty \{1-G_N(t)\}^{k-1} dG_{N-c_7}(t).$$

(B) PROCEDURES R_8 AND R_9 AND THEIR PROPERTIES

Let t_0 be a fixed positive number. Observe the number of arrivals $X_1(t_0),\ldots,X_k(t_0)$ from processes $\pi_1,\ldots,\pi_k$ during time t_0 respectively.

R_8: Select process π_i if and only if

$$(3.2.14)\qquad X_i(t_0)+1 \geq c_8 \max_{1\leq j\leq k} X_j(t_0)$$

where $c_8(\geq 0)$ is the maximum value for which the P*-condition is satisfied. It should be pointed out that the motivation of this type of rule lies in the result of Chapman [19] who showed that there is no unbiased estimator of the ratio $\frac{\theta_1}{\theta_2}$ with finite variance, where θ_1 and θ_2 are expected values of two independent Poisson distributions Y_1 and Y_2 respectively, but that the estimator $\frac{Y_1}{Y_2+1}$ is "almost unbiased".

Procedures R_8 was proposed by Gupta and Huang [39] in studying the single sample selection problem for Poisson populations. Here we derive some further properties of this selection rule. It is easy to see that for any $\underline{\lambda} = (\lambda_1,\ldots,\lambda_k) \in \Omega$,

$$(3.2.15)\quad p_{\underline{\lambda}}(i) = \sum_{x=0}^{\infty} e^{-\lambda_{[i]}t_0} \frac{(\lambda_{[i]}t_0)^x}{x!} \prod_{\substack{j=1\\ j\neq i}}^{k} \int_{\lambda_{[j]}t_0}^{\infty} \frac{1}{[\frac{x+1}{c_8}]!} y^{[\frac{x+1}{c_8}]} e^{-y} dy$$

where $[\frac{x+1}{c_8}]$ denotes the integer part of $\frac{x+1}{c_8}$.

It follows from (3.2.15) that R_8 is strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$. In particular,

$$(3.2.16)\quad \inf_{\Omega} P(CS|R_8) = \inf_{\alpha>0} \sum_{x=0}^{\infty} e^{-\alpha} \frac{\alpha^x}{x!} \{ \sum_{i=0}^{[\frac{1+x}{c_8}]} e^{-\alpha} \frac{\alpha^i}{i!}\}^{k-1}.$$

One can follow the method of Gupta and Huang to obtain a conservative value of the selection constant. For more detail, see [39].

Let $\Omega_1 = \{\underline{\lambda} = (\lambda,\ldots,\lambda,\delta\lambda): \lambda \geq \lambda_0 > 0\}$, $\delta > 1$. Then for any $\underline{\lambda} \in \Omega_1$,

$$E_{\underline{\lambda}}(S|R_8) = P_{\underline{\lambda}}(X_{(k)}(t_0)+1 \geq c_8 \max_{1\leq j\leq k-1} X_{(j)}(t_0)) + (k-1)P_{\underline{\lambda}}(X_{(1)}(t_0)+1 \geq c_8 \max X_{(j)}(t_0))$$

$$\leq k - P_{\underline{\lambda}}(X_{(k)}(t_0)+1 < c_8 X_{(1)}(t_0)) - (k-1)P_{\underline{\lambda}}(X_{(1)}(t_0)+1 < c_8 X_{(k)}(t_0))$$

$$= k - \sum_{x=2}^{\infty}\{\sum_{i=0}^{[\frac{c_8x+1}{1+c_8}]} \binom{x}{i}(\frac{\delta}{1+\delta})^i(\frac{1}{1+\delta})^{x-i}\}e^{-\lambda t_0(1+\delta)} \frac{(\lambda t_0(1+\delta))^x}{x!}$$

$$- (k-1)\sum_{x=0}^{\infty}\{\sum_{i=0}^{[\frac{c_8x+1}{1+c_8}]} \binom{x}{i}(\frac{1}{1+\delta})^{x-i}\}e^{-\lambda t_0(1+\delta)} \frac{(\lambda t_0(1+\delta))^x}{x!}$$

(3.2.17) $$\leq k-\{\inf_{x\geq 2} g(x)+(k-1)\inf_{x\geq 2} h(x)\}\{1-e^{-\lambda_0 t_0(1+\delta)}(\lambda_0 t_0(1+\delta))(1+\frac{1}{2}\lambda_0 t_0(1+\delta))\},$$

where g(x) and h(x) are defined in terms of incomplete beta function as follows:

$$g(x) = 1-I_{\frac{\delta}{1+\delta}}([\frac{c_8x-1}{1+c_8}]+1,\ x-[\frac{c_8x-1}{1+c_8}]),$$

(3.2.18)

$$h(x) = 1-I_{\frac{1}{1+\delta}}([\frac{c_8x-1}{1+c_8}]+1,\ x-[\frac{c_8x-1}{1+c_8}]).$$

Note that the upper bound (3.2.17) is better than the bound given in [39].

Using the same sampling rule as in R_8, the following conditional procedure was also proposed by Gupta and Huang [39] for the selection problem of k Poisson populations. Their procedure written in terms of observations in a fixed time to from each process, is

R_9: Select process π_i if and only if

(3.2.19) $$X_i(t_0)+1 \geq c_9 \max_{1\leq j\leq k} X_j(t_0) \quad \text{given} \sum_{i=1}^{k} X_i(t_0) = r,$$

where $c_9 \geq 0$ is the maximum value for which the P*-condition is satisfied.

Let $\underline{\lambda} = (\lambda_1,\ldots,\lambda_k) \in \Omega$, and let

$$s_i = \lambda_{[1]} +\ldots+ \lambda_{[i]}, \quad i \leq k,$$

(3.2.20)

$$p_{ij} = \frac{\lambda_{[j]}}{s_i}, \quad j = 1,\ldots,i.$$

Then

$$P_{\underline{\lambda}}(CS|R_9) = P_{\underline{\lambda}}(X_{(k)}(t_0)+1 \geq c_9 \max_{1\leq j\leq k-1} X_{(j)}(t_0) | \sum_{i=1}^{k} X_{(i)}(t_0)=r)$$

(3.2.21)
$$= \sum_{x=0}^{r} \binom{r}{x} p_{kk}^{x}(1-p_{kk})^{r-x} \sum (r-x)! \prod_{j=1}^{k-1} \left(\frac{p_{k-1,j}^{x_j}}{x_j!}\right)$$

where the second summation of the right hand side of (3.2.21) is over all (k-1)-tuples $(x_1,\ldots,x_{k-1})$ of nonnegative integers, such that $0 \leq x_i \leq \min\{\frac{x+1}{c_9}, r-x\}$, $i = 1,\ldots,k-1$, and $\sum_{i=1}^{k-1} x_i = r-x$.

Recall that the vector $\underline{x} = (x_1,\ldots,x_n)$ majorizes the vector $\underline{y} = (y_1,\ldots,y_n)$ if

$$\sum_{i=1}^{m} x_{[n+1-i]} \geq \sum_{i=1}^{m} y_{[n+1-i]} \quad \text{for } m = 1,\ldots,n-1, \quad \text{and}$$

(3.2.22)

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]},$$

and is written $\underline{x} > \underline{y}$. A real-valued function $\varphi(\underline{x})$ is called a Schur-convex (concave) function if $\varphi(\underline{x}) \geq (\leq)\varphi(\underline{y})$ whenever $\underline{x} > \underline{y}$. It is known that (see Rinott [79]) if $\varphi(x_1,\ldots,x_k)$ is a symmetric Schur-concave function and $(X_1,\ldots,X_k)$ is a multinomial random vector with parameter N and $\underline{p}$, then $E_{\underline{p}}\{\varphi(X_1,\ldots,X_k)\}$ is Schur-concave in $\underline{p}$.

Now for a fixed x, the second summation of the right hand member of (3.2.21) can be expressed as

$$(3.2.23)\quad \sum(r-x)!\prod_{j=1}^{k-1}\left(\frac{p_{k-1,j}^{x_j}}{x_j!}\right) = E_{\underline{p}}\{\psi(Y_1,\ldots,Y_{k-1})\}$$

where

$$(3.2.24)\quad \psi(y_1,\ldots,y_{k-1}) = \begin{cases} 1 & \text{if } c_3 \max\limits_{1\le j\le k-1} y_j \le x+1 \\ 0 & \text{if } c_3 \max\limits_{1\le j\le k-1} y_j > x+1, \end{cases}$$

and $(Y_1,\ldots,Y_{k-1})$ is a multinomial random vector with parameters r-x and $\underline{p} = (p_{k-1,1},\ldots,p_{k-1,k-1})$. Since ψ is a symmetric Schur-concave function, it follows that $E_{\underline{p}}\{\psi(Y_1,\ldots,Y_{k-1})\}$ is Schur-concave in $\underline{p}$. In other words, if we fix s_{k-1} and $\lambda_{[k]}$, then $P_{\underline{\lambda}}(CS|R_9)$ decreases when $\lambda_{[k-1]} \to \lambda_{[k]}$. Hence the least favorable configuration is of the form $(0,\ldots,0,\lambda,\delta\lambda,\ldots,\delta\lambda)$ where $\lambda > 0$, $\delta > 1$. It should be pointed out that the probability of a correct selection under the configuration $(0,\ldots,0,\lambda,\delta\lambda,\ldots,\delta\lambda)$ does not depend on the unknown parameter λ. Also when k = 2, the infimum takes place when the two processes are identical. However, when $k \ge 3$, the infimum of $P(CS|R_9)$ does not necessarily take place at the configuration of the type $(0,\ldots,0,\lambda,\ldots,\lambda)$ as shown by the following example.

First of all, we need some algebraic concepts. Let

(3.2.25) $P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$

be a polynomial of degree n. The coefficients $a_0, a_1, \ldots, a_n$ are assumed to be real. The discriminant of the polynomial p(x) is defined to be

$$(3.2.26)\quad D(p) = \begin{vmatrix} a_0 & a_1 \ldots a_{n-1} & a_n & 0 & 0 \ldots 0 \\ 0 & a_0 \ldots & a_{n-1} & a_n & 0 \ldots 0 \\ \vdots & & \vdots & \vdots & \vdots \quad \vdots \\ 0 & 0 \ \ldots \ \ldots & \ldots & 0 & a_0 \ldots a_n \\ na_0 & (n-1)a_1 \ldots a_{n-1} & 0 & 0 & 0 \ldots 0 \\ 0 & na_0 \ldots & a_{n-1} & 0 & 0 \ldots 0 \\ \vdots & & na_1 & & \ldots a_{n-1} \end{vmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} (n-1) \text{ rows} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} n \text{ rows} \end{matrix}$$

In particular, when $p(x) = a_0x^2 + a_1x + a_2$, then $D(p) = a_1^2 - 4a_0a_2$. It is well-known that (see [65]) if the polynomial p(x) with real coefficients, not having multiple roots, then $D(p) > 0$, if the number of pairs of complex conjugate roots of p(x) is even, and $D(p) < 0$, if this number is odd. Moreover, $D(p) = 0$ if and only if p(x) has multiple roots. Now we consider the case when $k = 3$, $c = \frac{2}{9}$ and $\sum_{i=1}^{3} X_{(i)}(t_0) = 6$ under the configuration

(3.2.27) $\quad (\lambda, \delta\lambda, \delta\lambda), \quad \lambda > 0, \ \delta \geq 1.$

Let $x = (1+2\delta)^{-1}$. Under the configuration (3.2.27),

$$\text{(3.2.28)}\quad P(CS|R_9) = P_r\Big(X_{(3)}(t_0)+1 \geq \frac{2}{9}\max_{1\leq i\leq 2} X_{(i)}(t_0)\,\Big|\,\sum_{i=1}^{3} X_{(i)}(t_0)=6\Big)$$

$$= 1-P_r\Big(X_{(3)}(t_0)=0,\ \max_{1\leq i\leq 2} X_{(i)}(t_0) \geq 5\,\Big|\,\sum_{i=1}^{3} X_{(i)}(t_0)=6\Big)$$

$$= 1-x^5(3-2x) - \Big(\frac{1-x}{2}\Big)^5\Big(\frac{1+11x}{2}\Big)$$

$$= p(x),\ \text{say.}$$

The derivative $p'(x)$ of $p(x)$ is a polynomial of degree 5. It follows that $p'(x)$ has at most two pairs of complex conjugate roots. Direct computation shows that the discriminant of $p'(x)$ is negative. This implies that $p'(x)$ has three real roots. Since $p'(\frac{1}{3})p'(1) < 0$ and $p'(x) \leq 0$ for all $0 < x < \frac{1}{11}$, there are at most two real roots lying in the interval $(\frac{1}{11}, \frac{1}{3})$. Moreover $p'(\frac{33}{352}) < 0 < p'(\frac{33}{352} + \frac{1}{11264})$, and $p'(\frac{1}{4}) > 0 > p'(\frac{1}{3})$. Now

$$P(\tfrac{33}{352}) = 0.980578 < p(1) = 0.9805795 < p(0) = 0.984375.$$

This implies that the infimum of $p(x)$ takes place for some x in the closed interval $[\frac{33}{352}, \frac{33}{352} + \frac{1}{11264}]$. This shows that the least favorable configuration is of the type $(\lambda,\delta\lambda,\delta\lambda)$, where $1 < \delta$.

The next theorem shows that R_9 is a monotone procedure.

<u>Theorem 3.2.3</u>. Procedure R_9 is monotone.

<u>Proof</u>. Let $1 \leq i < j \leq k$.

(3.2.29) $P(\pi_{(i)} \text{ is selected} | R_9)$

$$= P_r(X_{(i)}(t_0)+1 \geq c_9 \max X_{(\ell)}(t_0) | \sum_{\ell=1}^{k} X_{(\ell)}(t_0) = r)$$

$$= \sum \binom{r}{x_i+x_j, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k} (p_{[i]}+p_{[j]})^{x_i+x_j} \prod_{\substack{\ell=1 \\ \ell \neq i \\ \ell \neq j}}^{k} p_{[\ell]}^{x_\ell} \cdot$$

$$I_{\frac{p_{[i]}}{p_{[i]}+p_{[j]}}}(s, x_i+x_j-s+1),$$

where $s = \max\{(1+c_9)^{-1}\{c_9(x_i+x_j)-1\},\ c_9 \max_{\substack{\ell \neq i \\ \ell \neq j}} x_\ell - 1\}$ and

$I_p(\cdot)$ represents the incomplete beta function, and the summation extends to all k-tuples $(x_1, \ldots, x_k)$ of nonnegative integers such that $\sum_{i=1}^{k} x_i = r$ and $x_i \geq s$. A similar expression for the probability that $\pi_{(j)}$ is included in the selected subset can be obtained by interchanging the roles of x_i and x_j in (3.2.29). From the fact that

$$I_{\frac{p_{[i]}}{p_{[i]}+p_{[j]}}}(s, x_i+x_j-s+1) \leq I_{\frac{p_{[j]}}{p_{[i]}+p_{[j]}}}(s, x_i+x_j-s+1),$$

the result follows.

<u>Theorem 3.2.4</u>. For a given P^*, let $P_1^* = \frac{1-P^*}{k-1}$ and let c_9 be the largest positive number such that $\sum_{i=0}^{[\frac{c_9 r-1}{1+c_9}]} \binom{r}{i} \frac{1}{2^r} \leq P_1^*$, then $\inf_{\Omega} P(CS|R_9) \geq P^*$.

Proof. For any $\underline{\lambda} = (\lambda_1,\ldots,\lambda_k) \in \Omega$,

$$P_{\underline{\lambda}}(CS|R_9) = P_r(X_{(k)}(t_0)+1 \geq c_9 \max_{1\leq j\leq k-1} X_{(j)}(t_0) | \sum_{i=1}^{k} X_{(i)}(t_0)=r)$$

$$= 1-P_r(X_{(k)}(t_0)+1 < c_9 \max_{1\leq j\leq k-1} X_{(j)}(t_0) | \sum_{i=1}^{k} X_{(i)}(t_0)=r)$$

$$\geq 1-\sum_{j=1}^{k-1} P_r(X_{(k)}(t_0)+1 < c_9 X_{(j)}(t_0) | \sum_{i=1}^{k} X_{(i)}(t_0)=r)$$

$$\geq 1-\sum_{j=1}^{k-1} \{ \sum_{i=0}^{[\frac{c_9 r-1}{1+c_9}]} \binom{r}{i} \left(\frac{\lambda_{[k]}}{\lambda_{[k]}+\lambda_{[j]}}\right)^i \left(\frac{\lambda_{[j]}}{\lambda_{[k]}+\lambda_{[j]}}\right)^{r-i} \}$$

$$\geq 1-(k-1) \sum_{i=0}^{[\frac{c_9 r-1}{1+c_9}]} \binom{r}{i} \frac{1}{2^r}$$

$$= P^*.$$

(C) PROCEDURE R_{10} AND ITS PROPERTIES

Suppose that the Poisson processes are observed at successive intervals of time, $t = 1,2,\ldots$. Observe the processes until time t_0, the smallest value of t, say, when the number of arrivals from one of the processes is equal to or greater than N. Let I denote the set of values of i for which $X_i(t_0) \geq N$ and J the set of values j for which $X_j(t_0) \geq N-c_7$ where c_7 is the constant associated with R_7 defined in (3.2.1). Clearly $I \subseteq J$. For each $j \in J$, let t_{j_0} be the time such that $X_j(t_{j_0}) \geq N-c_7$ and $X_j(t_{j_0}-1) < N-c_7$, and let $m_j = N-c_7-X_j(t_{j_0}-1)$, $n_j = X_j(t_{j_0})-X_j(t_{j_0}-1)$. Similarly for each $i \in I$, let $m_i' = N-X_i(t_0-1)$ and $n_i' = X_i(t_0)-X_i(t_0-1)$. Let U(m,n) denote the mth smallest observation in a sample of size n from a uniform distribution on the unit interval (0,1). Now we compute

$$(3.2.30)\qquad \begin{aligned} U_j &= t_{j_0}-1 + U(m_j,n_j) && \text{for } j \in J \\ U_i' &= t_0-1 + U(m_i',n_i') && \text{for } i \in I \end{aligned}$$

and propose the following selection rule:

R_{10}: Select process π_j ($j \in J$) if and only if

$$(3.2.31)\qquad U_j \leq \min_{i \in I} U_i'$$

Note that U_i' and U_j are simply the waiting times for N and $N-c_7$ arrivals from the processes π_i and π_j, respectively. To see this, observe that if n is a random variable distributed according to the Poisson distribution with mean λ, then for any given value of m,

$$(3.2.32)\qquad P_r(U(m,n) \leq t) = \sum_{n=m}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \cdot \frac{n!}{(m-1)!(n-m)!} \int_0^t x^{m-1}(1-x)^{n-m}dx$$

$$= G_m(\lambda t), \qquad 0 < t < 1.$$

Thus, the arrival times for a Poisson process can be generated from the observed number of arrivals during the successive unit time intervals and random observations from a uniform distribution. It follows that for any $\underline{\lambda} \in \Omega$,

$$P_{\underline{\lambda}}(\pi_{(i)} \text{ is selected} | R_{10}) = P_{\underline{\lambda}}(\pi_{(i)} \text{ is selected} | R_7).$$

Hence the rule R_{10} is strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$.

In particular,

$$(3.2.33) \qquad \inf_{\Omega} P(CS|R_{10}) = \int_0^{\infty}\{1-G_N(t)\}^{k-1} dG_{N-c_7}(t).$$

$$= \frac{1}{k} \sup_{\Omega} E(S|R_{10}).$$

3.3 SELECTION PROCEDURES FOR THE PROCESS ASSOCIATED WITH $\lambda_{[1]}$

For the analogous problem of selecting the process for which the mean waiting time is largest, we propose the following subset selection rule.

(A) Let N be a fixed positive integer. We observed the processes until, say to, that $\min_{1 \le i \le k} X_i(t_0) = N$.

R_{11}: Select the process π_i if and only if

$$(3.3.1) \qquad X_i(t_0) \le N+c_{11},$$

where c_{11} is the smallest non-negative integer such that the basic probability requirement is satisfied.

By using similar arguments as given in Section 2, one can show that the procedure R_{11} is reverse strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$. This implies that the infimum of $P(CS|R_{11})$ takes place when all the process are identical. In fact, the infimum of $P(CS|R_{11})$ is given by

$$(3.3.2) \qquad \inf_{\Omega} P(CS|R_{11}) = \int_0^{\infty} G_N^{k-1}(t) dG_{N+c_{11}}(t).$$

Also, one can show that

(3.3.3) $$\sup_{\Omega} E(S|R_{11}) = k \inf_{\Omega} P(CS|R_{11}).$$

(B) Suppose we use the same stopping rule as mentioned in (A), we can define a selection rule as follows:

R_{12}: Select process π_i if and only if

(3.3.4) $$T_i(N) \geq c_{12} \max_{1 \leq j \leq k} T_j(N)$$

where $0 < c_{12} \leq 1$ is chosen so that the basic probability requirement holds.

Note that this reduces to the problem of selecting a subset of k gamma populations which includes the one with the smallest value of shape parameter. We briefly state below some known results. For more details, one can see [34].

(i) Rule R_{12} is reverse strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$.

(ii) $\sup_{\Omega} E(S|R_{12}) = k \inf_{\Omega} P(CS|R_{12})$

(C) If the processes are observed at successive intervals of time $t = 1,2,\ldots$. We observe the processes until the first time to, say, when $\min_{1 \leq i \leq k} X_i(t_0) \geq N$. Let t_i be the time such that $X_i(t_i) \geq N$ and $X_i(t_i-1) < N$, $i = 1,\ldots,k$. Let $m_i = N-X_i(t_i-1)$ and $n_i = X_i(t_i) - X_i(t_i-1)$. As in the previous section, we compute

(3.3.5) $U_i = t_i - 1 + U(m_i, n_i), \quad i = 1,\ldots,k,$

and propose a procedure R_{13} as follows:

R_{13}: Retain process π_i in the selected subset if and only if

(3.3.6) $$U_i \geq c_{13} \max_{1 \leq j \leq k} U_j,$$

where $0 < c_{13} \leq 1$ is the largest value for which the p*-condition is satisfied. Since U_i is distributed as $T_i(N)$, hence the rule R_{13} is equivalent to R_{12}. It should be pointed out that a rule similar to R_{13} has been studied by Goel [29].

3.4 APPLICATIONS

(A) A sequential (inverse sampling) subset selection rule for the most probable multinomial event.

Let $\underline{X} = (X_1,\ldots,X_k)$ have the multinomial distribution

(3.4.1) $$P(\underline{X} = \underline{x}) = \binom{n}{x_1,\ldots,x_m} \prod_{i=1}^{k} p_i^{x_i}$$

where $\underline{x} = (x_1,\ldots,x_m)$, and let $p_{[1]} \leq \cdots \leq p_{[k]}$ denote the ordered values of $p_1,\ldots,p_k$. The subset selection problem for the multinomial distribution has been considered by Gupta and Nagel [44], Gupta and Huang [39] and Panchapakesan [70]. A related problem has also been discussed by Alam, Seo and Thompson [2], Bechhofer, Elmarghrabi and Morse [12]. In [39] and [44], the authors considered the single-sample subset selection rules. The procedure given in [70] is based on a completely sequential sampling scheme in which one observation is taken at a time from the given distribution until the highest cell count is equal to a fixed number N, say.

We consider below a variation of the sampling scheme in [2]. Our sampling scheme is given as follows: Let a positive integer N be given, and let $n_1, n_2, \ldots$ denote a sequence of random observation taken from a Poisson distribution with mean λ. Having observed these number, taken n_i observations from the given multinomial distribution for the i<u>th</u> experiment, $i = 1,2,\ldots$. Let π_i denote the cell corresponding to p_i, and let Y_{ij} denote the cell count in π_i out of n_j observations. Stop sampling as soon as the total count from any cell is equal to or greater than N. Let t_0 denote the stage at which the experiment terminates, and let $X_i(t) = \sum_{j=1}^{t} Y_{ij}$. Then $X_i(t_0-1) < N$ for $i = 1,\ldots,k$ and $X_i(t_0) \geq N$ for some i. As in Section 3.2, let I be the set of values of i for which $X_i(t_0) \geq N$ and J be the set of values of j for which $X_j(t_0) \geq N-c_7$ where c_7 is the selection constant associated with rule R_7. Take the similar random observations from the uniform distribution on the unit interval (0,1) and obtain the statistics U_i' and U_j as defined in (3.2.30). Based on the statistics U_i''s and U_j's, we select the cell according to the rule R_{10}. Then the problem reduces to that of selecting the Poisson process with minimum mean waiting time. To see this, suppose the parameter n in (3.4.1) is a random variable distributed according to a Poisson distribution with mean λ. It is easy to show that the cells frequencies $X_1,\ldots,X_k$ are independently distributed according to the Poisson distribution with mean $\lambda p_1,\ldots,\lambda p_k$, respectively. It follows from Theorem 3.2.1 that the least favorable configuration is $(\frac{\lambda}{k},\ldots,\frac{\lambda}{k})$, and the infimum of P(CS) is independent on the parameter λ. Moreover supremum of the expected subset size is obtained when all the cells are identical and is equal to k $\inf_{\Omega}$ P(CS). It should be pointed out that when $\lambda \to 0$, the rule reduces to the one proposed by Panchapakesan [70].

(B) A sequential (inverse sampling) rule for selection procedure for k binomial populations

Let $\pi_1,\ldots,\pi_k$ be k independent binomial populations with parameters $p_1,\ldots,p_k$ respectively. To select a subset of the k populations which contains the population associated with the largest p_i, Gupta and Sobel [48] proposed a single-sample procedure which is based on the statistics $\max_{1\le j\le k} X_j-X_i$, where X_i represents the number of successes in n independent trials from populations π_i. Recently, Gupta, Huang and Huang [41] proposed a conditional procedure for this problem and gave a lower bound for the infimum of the probability of a correct selection. It should be pointed out that a related problem has been considered by Sobel and Weiss [91].

Suppose the number of observations taken at each stage from the k binomial populations, is a random variable distributed according to a Poisson distribution with mean λ. Using the same sampling procedure and selection rule as mentioned in part (A) of this section, the problem then reduces to that of selecting the Poisson process with smallest mean waiting time. It follows that the infimum of the probability of a correct selection and the supremum of the expected subset size take place when all the populations are identical. Also the inf P(CS) and the sup E(S) do not depend on the common unknown parameter p and the mean λ. Moreover, the selection rule is strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$.

CHAPTER IV

SUBSET SELECTION PROCEDURES FOR FINITE SCHEMES IN INFORMATION THEORY

4.1 INTRODUCTION

In probability theory a complete system of events $A_1,\ldots,A_m$ means a set of events such that one and only one of them must occur at each trial (e.g., the appearance of 1,2,3,4,5 or 6 in a throw of a die). In the case m = 2, we have a pair of mutually exclusive events. If we are given the events $A_1,A_2,\ldots,A_m$ of a complete system, together with their probabilities $p_1,p_2,\ldots,p_m$ ($p_i \geq 0$, $\sum_{i=1}^{m} p_i = 1$), then we say that we have a finite scheme

$$\pi = \begin{pmatrix} A_1 & A_2 & \cdots & A_m \\ p_1 & p_2 & \cdots & p_m \end{pmatrix}. \tag{4.1.1}$$

Every finite scheme describes a state of uncertainty. We have an experiment, the outcomes of which must be one of the events $A_1,A_2,\ldots,A_m$, and we know only the probabilities of these outcomes. It seems obvious that the amount of uncertainty is different in different schemes. Thus in the two simple alternatives (see Khinchin [57]),

$$\begin{pmatrix} A_1 & A_2 \\ 0.5 & 0.5 \end{pmatrix}, \qquad \begin{pmatrix} A_1 & A_2 \\ 0.99 & 0.01 \end{pmatrix},$$

the first obviously represents much more uncertainty than the second; in the second case, the result of the experiment is "almost sure" A_1, while in the first case we naturally refrain from making any predictions. The scheme

$$\begin{pmatrix} A_1 & A_2 \\ 0.3 & 0.7 \end{pmatrix}$$

represents an amount of uncertainty intermediate between the proceeding two.

For many applications it seems desirable to introduce an ordering so as to compare the amount of uncertainty between two different finite schemes. For example, Shannon's entropy function

$$(4.1.2) \qquad H(p_1,\ldots,p_m) = -\sum_{i=1}^{m} p_i \log p_i$$

can serve as a measure of the uncertainty of the finite scheme (4.1.1); (we always take $p \log p = 0$ if $p = 0$). For general properties of entropy function, see Shannon [87] and Feinstein [25]. In one of his several papers, Rényi [77] obtained a characterization of Shannon's measure of entropy for generalized probability distributions. Vincze [98] and Perez [72] have discussed some problems of statistical decision theory from the point of view of information theory. For results which simplify and extend Shannon's original work, see Csiszár [20], Arimoto [3], Kendall [56] and Lee [59]. The applications of information-theoretic concepts to physics, chemistry, and various branches of biology and psychology have been discussed by Brillouin [17] and Quastler [75], [76]. In [38], Gupta and Huang have investigated a selection procedure for the entropy function associated with the binomial populations. Gupta and Huang studied the subset selection problem for the selection of the population associated with the largest unknown entropy function which is equivalent to the selection of the binomial population associated with $\min_{1\leq i\leq k} |p_i - \frac{1}{2}|$. They also proposed a general selection procedure by using the generalized entropy function introduced by Arimoto [3].

In this chapter, we partially order the parameter space by means of majorization and propose a subset selection rule based on a Schur-concave function. In Section 4.2, some properties of the selection procedure are discussed. The infimum of the probability of a correct selection is shown to occur when the k given finite schemes are identical. A method leading to a conservative solution for the selection constant is also given. An upper bound for the expected subset size is given in Section 4.5. In Section 4.7, we discuss an application to testing of homogeneity. Some related problems are also discussed in Section 4.8.

4.2 SELECTION PROCEDURES

First of all, let us recall the definition of majorization. The vector $\underline{x} = (x_1,\ldots,x_m)$ is said to majorize the vector $\underline{y} = (y_1,\ldots,y_m)$ if

$$\sum_{i=1}^{r} x_{[m+1-i]} \geq \sum_{j=1}^{r} y_{[m+1-j]}, \quad \text{for } r = 1,\ldots,m-1, \text{ and}$$

$$\sum_{i=1}^{m} x_{[i]} = \sum_{i=1}^{m} y_{[i]}.$$

In this case, we denote $\underline{x} \succ \underline{y}$ or, equivalently, $\underline{y} \prec \underline{x}$.

Note that in general this is only a partial order, but when $m = 2$, it is a linear order defined on the simplex $S = \{(a_1,a_2): a_1 + a_2 = \text{constant}\}$. For if $\underline{a} = (a_1,a_2)$ and $\underline{b} = (b_1,b_2)$ are two points in S with $a_1 \geq a_2$, $b_1 \geq b_2$ then $\underline{a} \succ \underline{b}$ if and only if $a_1 \geq b_1$.

Also recall that if a function φ satisfies the property that $\varphi(\underline{x}) \leq \varphi(\underline{y})$ ($\varphi(\underline{x}) \geq \varphi(\underline{y})$) whenever $\underline{x} \succ \underline{y}$, then φ is called a Schur-concave (Schur-convex) function. Functions which are either Schur-concave or Schur-convex are called Schur functions. Note that a Schur function is necessarily permutation invariant; that is $\varphi(\underline{x}) = \varphi(\underline{x}')$ whenever $\underline{x}' = (x_{\alpha_1},\ldots,x_{\alpha_m})$ and $(\alpha_1,\ldots,\alpha_m)$ is a permutation of $(1,\ldots,m)$. Furthermore, all functions of the form $\sum_{i=1}^{m} g(x_i)$, with g concave (convex), are Schur-concave (Schur-convex) functions. In particular, $H(p_1,\ldots,p_m)$ is a Schur-concave function.

Now let $\pi_i = \begin{pmatrix} A_1 & \cdots & A_m \\ p_{i1} & \cdots & p_{im} \end{pmatrix}$, $i = 1,\ldots,k$, be k independent finite schemes where each $p_{ij} \geq 0$, $\sum_{j=1}^{m} p_{ij} = 1$, $i = 1,\ldots,k$. We say that scheme π_i is better than scheme π_j if $\underline{p}_i = (p_{i1},\ldots,p_{im}) \prec \underline{p}_j = (p_{j1},\ldots,p_{jm})$. Since entropy function H is Schur-concave, hence $\underset{\sim}{p}_i \prec \underset{\sim}{p}_j$ implies that $H(\underset{\sim}{p}_i) \geq H(\underset{\sim}{p}_j)$. However, one entropy may arise from two different or non-comparable vectors. This is easily seen from the following argument: the entropy of $(\frac{1}{2}, \frac{1}{2}, 0)$ is log 2. If we consider vectors of the form $(x, \frac{1-x}{2}, \frac{1-x}{2})$ with $\frac{1}{2} \leq x \leq 1$, then since $H(x, \frac{1-x}{2}, \frac{1-x}{2})$ is continuous in x and $H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) > \log 2 > H(1,0,0) = 0$, hence there exists a number x, $\frac{1}{2} < x < 1$ such that $H(x, \frac{1-x}{2}, \frac{1-x}{2}) = H(\frac{1}{2}, \frac{1}{2}, 0)$. Clearly vectors $(x, \frac{1-x}{2}, \frac{1-x}{2})$ and $(\frac{1}{2}, \frac{1}{2}, 0)$ are noncomparable in the sense that one does not majorize the other. Therefore if we know only the entropy of the scheme (4.1.1) and we take n independent observations from this scheme, we are still not able to compute the probability of an event say, $A_i = a_i$, $a_i \geq 0$, $i = 1,\ldots,m$ and $\sum_{i=1}^{m} a_i = n$. Bearing this in mind, it is assumed that among the k given schemes, there always exists one scheme which is better than the others. This is equivalent to saying that there exists some i such that $\underset{\sim}{p}_i \prec \underset{\sim}{p}_j$ for all j, $1 \leq j \leq k$. This population is called the best population. If there are more than one "best" population, then we assume that one of them is tagged as the best. Our goal is to select a nonempty subset if the k schemes with the guarantee that the basic probability requirement is satisfied.

Suppose now we take n independent observations from each of the k schemes. Let X_{ij} denote the number of outcomes of j<u>th</u> event in scheme π_i. We propose a class of selection procedures R_φ as follows.

R_φ: Select scheme π_i if and only if

$$\varphi(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}) \geq \max_{1\leq j\leq k} \varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}) - d \qquad (4.2.1)$$

where φ is a Schur-concave function. Further, the selection constant d is chosen to be the smallest non-negative number such that the basic probability requirement is satisfied. Note that when m = 2 and φ is taken to be the Shannon's entropy function, (4.2.1) reduces to the procedure proposed by Gupta and Huang [38]. Furthermore, besides the work of Gupta and Huang [38], very little literature on ranking and selection problems in information theory is available. In this chapter, we concentrate on the cases $m \geq 3$ which are also more practical and useful.

4.3 SOME PROPERTIES OF THE CLASS OF SELECTION PROCEDURES R_φ

Let $\Omega = \{\underset{\sim}{\omega} = (\underset{\sim}{p}_1,\ldots,\underset{\sim}{p}_k)$: there exists some i such that $\underset{\sim}{p}_i \prec \underset{\sim}{p}_j \ \forall_j\}$, and let

$$(4.3.1) \quad P_{\underset{\sim}{\omega}}(i|R_\varphi) = P_{\underset{\sim}{\omega}}(\text{Select } \pi_i \text{ with } \underset{\sim}{p}_i|R_\varphi), \quad i = 1,\ldots,k.$$

Theorem 4.3.1. For any Schur-concave function φ, and $1 \leq i \leq k$,

(i) $P_{\underset{\sim}{\omega}}(i|R_\varphi)$ is a Schur-concave function of $\underset{\sim}{p}_i$ when all other components of $\underset{\sim}{\omega}$ are fixed.

(ii) $P_{\underset{\sim}{\omega}}(i|R_\varphi)$ is a Schur-convex function of $\underset{\sim}{p}_j$ $(j \neq i)$ when all other components of $\underset{\sim}{\omega}$ are fixed.

Proof. Since

$$(4.3.2) \quad P_{\underset{\sim}{\omega}}(i|R_\varphi) = P_{\underset{\sim}{\omega}}\left(\varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) \geq \max_{1\leq j\leq k} \varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right)-d\right)$$

$$= \sum \prod_{\substack{j=1\\ j\neq i}}^{k} P_{\underset{\sim}{p}_j}\left(\varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right) \leq \varphi\left(\frac{x_1}{n},\ldots,\frac{x_m}{n}\right)+d\right) \cdot$$

$$P_{\underset{\sim}{p}_i}((X_{i1},\ldots,X_{im}) = (x_1,\ldots,x_m))$$

where the summation is over all m-tuples of nonnegative integer such that $\sum_{i=1}^{m} x_i = n$. Note that (4.3.2) can be rewritten as follows:

$$(4.3.2) \quad P_{\underset{\sim}{\omega}}(i|R_\varphi) = E_{\underset{\sim}{p}_i}\{\phi_i(X_{i1},\ldots,X_{im})\}$$

$$= E_{\underset{\sim}{p}_i}\{\prod_{\substack{j=1\\j\neq i}}^{k} \psi_j(X_{i1},\ldots,X_{im})\}$$

where

$$(4.3.3) \quad \psi_j(x_1,\ldots,x_m) = E_{\underset{\sim}{p}_j}\{I_{\{\varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}) \leq \varphi(\frac{x_1}{n},\ldots,\frac{x_m}{n})+d\}}\}$$

where $(X_{\ell 1},\ldots,X_{\ell m})$ represents a multinomial random vector with parameters n and $\underset{\sim}{p}_\ell$, and I_A represents the indicator function of A. Since for any j ($j \neq i$), $\psi_j(x_1,\ldots,x_m)$ is a nonnegative Schur-concave function, it follows that $\prod_{\substack{j=1\\j\neq i}}^{k} \psi_j(x_1,\ldots,x_m)$ is Schur-concave. Hence $E_{\underset{\sim}{p}_i}\{\prod_{\substack{j=1\\j\neq i}}^{k} \psi_j(X_{i1},\ldots,X_{im})\}$ is Schur-concave in $\underset{\sim}{p}_i$. This proves part (i).

Part (ii) follows immediately from the fact that while keeping $(x_1,\ldots,x_m)$ fixed, $\psi_j(x_1,\ldots,x_m)$ is a Schur-convex function of $\underset{\sim}{p}_j$. This completes the proof.

<u>Corollary 4.3.1</u>. If $\underset{\sim}{p}_i \prec \underset{\sim}{p}_j$ then $P_{\underset{\sim}{\omega}}(i|R_\varphi) \geq P_{\underset{\sim}{\omega}}(j|R_\varphi)$.

<u>Proof</u>: Let $\underset{\sim}{\omega}^*$ be obtained from $\underset{\sim}{\omega}$ by replacing $\underset{\sim}{p}_j$ by $\underset{\sim}{p}_i$ and all other components of $\underset{\sim}{\omega}$ remain fixed. Then

$$P_{\underline{\omega}}(i|R_\varphi) \geq P_{\underline{\omega}^*}(i|R_\varphi) \quad \text{(by Theorem 4.3.1, (i)}$$
$$= P_{\underline{\omega}^*}(j|R_\varphi)$$
$$\geq P_{\underline{\omega}}(j|R_\varphi) \quad \text{(by Theorem 4.3.1, (ii)).}$$

By applying part (ii) of Theorem 4.3.1, it follows immediately that the worst configuration of the probability of a correct selection occurs when all the parameters are equal. In other words, we have

Corollary 4.3.2. $\inf_{\Omega} P(CS|R_\varphi) = \inf_{\Omega_0} P(CS|R_\varphi)$

where $\Omega_0 = \{\underline{\omega} = (\underline{p},\ldots,\underline{p}),\ \underline{p} = (p_1,\ldots,p_m),\ p_i \geq 0,\ i = 1,\ldots,k,$ and $\sum_{i=1}^{m} p_i = 1\}$.

Next let us show how to obtain a conservative value of the selection constant.

Let $\underline{t} = (t_1,\ldots,t_m)$, where $0 \leq t_i \leq kn$, $i = 1,\ldots,m$ and $\sum_{i=1}^{m} t_i = kn$, and let

$$(4.3.4) \qquad M(k,d(\underline{t}),\underline{t},m,n) = \sum \prod_{i=1}^{k} \binom{n}{s_{i1},\ldots,s_{im}}$$

where the summation is over the set of all m-tuples $(s_{i1},\ldots,s_{im})$ such that $0 \leq s_{ij} \leq n$, $i = 1,\ldots,k$, $j = 1,\ldots,m$, $\sum_{i=1}^{k} s_{ij} = t_j$, and

$$(4.3.5) \qquad \varphi\left(\frac{s_{k1}}{n},\ldots,\frac{s_{km}}{n}\right) \geq \max_{1 \leq j \leq k-1} \varphi\left(\frac{s_{j1}}{n},\ldots,\frac{s_{jm}}{n}\right) - d(\underline{t})$$

for some constant $d(\underline{t})$ depending on $\underline{t}$. It is easy to see the following lemma.

Lemma 4.3.1. For any $\underline{t} = (t_1,\ldots,t_m)$ with $0 \leq t_i \leq kn$, $i = 1,\ldots,m$ and $\sum_{i=1}^{m} t_i = kn$, let $\underline{\omega} \in \Omega_0$,

$$(4.3.6) \quad P_{\underline{\omega}}\left(\varphi\left(\frac{X_{k1}}{n},\ldots,\frac{X_{km}}{n}\right) \geq \max_{1\leq j\leq k-1} \varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right)-d(\underline{t}) \,\Big|\, \sum_{j=1}^{k} X_{ji} = t_i,\ i = 1,\ldots,m\right)$$

$$= \frac{M(k,d(\underline{t}),\underline{t},m,n)}{\binom{kn}{t_1 \ \ldots \ t_m}} .$$

Theorem 4.3.2. For given P* and for each $\underline{t} = (t_1,\ldots,t_m)$, $0 \leq t_i \leq kn$, $i = 1,\ldots,m$, $\sum_{i=1}^{m} t_i = kn$, let $d(\underline{t})$ be the smallest constant such that

$$(4.3.7) \quad M(k,d(\underline{t}),\underline{t},m,n) \geq \binom{kn}{t_1 \ \ldots \ t_m} P^*.$$

If $d = \max_{\underline{t}}\{d(\underline{t})\}$, then

$$\inf_{\Omega} P(CS|R_{\varphi}) \geq P^*.$$

Proof.

$$P_{\Omega_0}(CS|R_{\varphi}) = P_{\Omega_0}\left(\varphi\left(\frac{X_{k1}}{n},\ldots,\frac{X_{km}}{n}\right) \geq \max_{1\leq j\leq k-1} \varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right)-d\right)$$

$$= \sum P_{\Omega_0}\left(\varphi\left(\frac{X_{k1}}{n},\ldots,\frac{X_{km}}{n}\right) \geq \max_{1\leq j\leq k-1} \varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right)-d \,\Big|\, \sum_{j=1}^{k} X_{ji}=t_i,\ i = 1,\ldots,m\right)$$

$$\cdot P\left(\sum_{j=1}^{k} X_{ji} = t_i,\ i = 1,\ldots,m\right)$$

$$(4.3.8) \quad \geq \sum P_{\Omega_0}\left(\varphi\left(\frac{X_{k1}}{n},\ldots,\frac{X_{km}}{n}\right) \geq \max_{1\leq j\leq k-1} \varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right)-d(\underline{t}) \,\Big|\, \sum_{j=1}^{k} X_{ji}=t_i,\ i = 1,\ldots,m\right) P\left(\sum_{j=1}^{k} X_{ji} = t_i,\ i = 1,\ldots,m\right)$$

where the summation is over the same set of all m-tuples of nonnegative integers as defined in (4.3.5). By (4.3.8) and Corollary 4.3.2, the proof is complete.

For m = 2, and if φ is taken to be the entropy function, some tables for the selection constants are available in [38]. For the functions $\varphi(x) = x(1-x)$ and $\varphi(x) = \min(x,1-x)$, tables for the selection constants are given at the end of this chapter.

4.4 SPECIAL CASE, m = 2

In this section, we are going to make further investigation of the selection procedure R_φ. Since when m = 2, we know that $\underline{p}_i \prec \underline{p}_2$ if and only if $\varphi(\underline{p}_1) \geq \varphi(\underline{p}_2)$. Also in this case, the ordering becomes linear. Hence to order the schemes in terms of majorization, we may order them by means of $\varphi(\underline{p})$. Furthermore, the distribution of $\varphi(\frac{X}{n}, \frac{n-X}{n})$ depends on the parameter only through $\varphi(p,1-p)$. For simplicity, we write $\varphi(p,1-p) = \varphi(p)$ and $\varphi(\frac{X}{n}, \frac{n-X}{n}) = \varphi(\frac{X}{n})$. First of all, we prove the following theorem.

Theorem 4.4.1. Let $(X_1,\ldots,X_m)$ have multinomial distribution with parameter n and $\underline{p}$. For any constant c, $P_{\underline{p}}(\varphi(\frac{X_1}{n},\ldots,\frac{X_m}{n}) \geq c)$ is Schur-concave in $\underline{p}$.

Proof. Define

$$(4.4.1)\quad \phi_c(x_1,\ldots,x_m) = \begin{cases} 1 & \text{if } \varphi(\frac{x_1}{n},\ldots,\frac{x_m}{n}) \geq c \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that ϕ_c is a Schur-concave function. It follows that $E_{\underline{p}}\{\phi_c(X_1,\ldots,X_m)\}$ is Schur-concave in $\underline{p}$. Since $P\{\varphi(\frac{X_1}{n},\ldots,\frac{X_m}{n}) \geq c\} = E_{\underline{p}}\{\phi_c(X_1,\ldots,X_m)\}$. The proof is complete.

Corollary 4.4.1. If X is binomial random variable with parameters n and p, then for any constant c, $P\{\varphi(\frac{X}{n}) \geq c\}$ is nondecreasing in θ.

Let $f_\theta(x)$ be the density of $\varphi(\frac{X}{n})$ with $\theta = \varphi(p)$. The following theorem shows that the family $\{f_\theta\}$ has a monotone likelihood ratio in x. To show this, we prove, first of all, the following lemma.

Lemma 4.4.1. Let $0 \leq p_1 \leq p_2 \leq \frac{1}{2}$, and let $A_i(x) = p_i^x(1-p_i)^{n-x} + p_i^{n-x}(1-p_i)^x$, $i = 1,2$. Then $\frac{A_2(x)}{A_1(x)}$ is increasing in x for $x \in (0, \frac{n}{2})$.

Proof. For any $x \in (0, \frac{n}{2})$

$$\frac{d}{dx}\log\{\frac{A_2(x)}{A_1(x)}\} = \frac{1}{A_1(x)A_2(x)}\{p_1p_2(1-p_1)(1-p_2)\}^x\{((1-p_1)(1-p_2))^{n-2x} - (p_1p_2)^{n-2x}\}\log\frac{p_2(1-p_1)}{p_1(1-p_2)}$$

$$+ \frac{1}{A_1(x)A_2(x)}\{p_1p_2(1-p_1)(1-p_2)\}^x\{(p_2(1-p_1))^{n-2x} - (p_1(1-p_2))^{n-2x}\}\log\frac{(1-p_1)(1-p_2)}{p_1p_2}$$

$$\geq 0.$$

Hence the result follows.

Theorem 4.4.2. The family $\{f_\theta\}$ has a monotone likelihood ratio in x.

Proof. Suppose $\theta_1 = \varphi(p_1) \leq \varphi(p_2) = \theta_2$. Since $P_p\{\varphi(\frac{X}{n}) = c\} = P_{1-p}\{\varphi(\frac{X}{n}) = c\}$, hence we may assume that $0 \leq p_1 \leq p_2 \leq \frac{1}{2}$. For any $0 \leq y \leq \varphi([\frac{n}{2}])$, let x be the number such that $0 \leq x \leq \frac{n}{2}$ and $\varphi(\frac{X}{n}) = y$. Then

$$(4.4.2) \quad \frac{f_{\theta_2}(y)}{f_{\theta_1}(y)} = \frac{\binom{n}{x}\{p_2^x(1-p_2)^{n-x} + p_2^{n-x}(1-p_2)^x\}}{\binom{n}{x}\{p_1^x(1-p_1)^{n-x} + p_1^{n-x}(1-p_1)^x\}} = \frac{A_2(x)}{A_1(x)} .$$

By (4.4.2) and Lemma 4.4.1, the result follows.

Let us discuss the problem related to the probability of complete ranking of $k(k \geq 2)$ populations. To do this, we show first of all the following result.

Lemma 4.4.2. If $X_1,\ldots,X_k$ are k independently distributed random variables with parameter $\theta_1,\ldots,\theta_k$ respectively and with densities having monotone likelihood ratio property, and if a and b are two (extended) real numbers such that $a < b$, then

$$(4.4.3) \quad P_r\{a < X_1 \leq\cdots\leq X_i \leq X_{i+1}\leq\cdots\leq X_k<b\}\leq P_r\{a < X_1 \leq\cdots\leq X_i \leq X_{i+1} \leq\cdots\leq X_k < b\},$$

provided that $\theta_i \leq \theta_{i+1}$.

Proof. Let $g_i(x)$ be the density (probability mass function) of X_i, and let

$$A_i(x) = P_r\{a < X_1 \leq \ldots \leq X_{i-2} \leq x\}$$

(4.4.4) $$B_i(x) = P_r\{x \leq X_i \leq x_{i+1} \leq \ldots \leq X_k < b\}$$

$$B_i^*(x) = P_r\{x \leq X_{i+1} \leq X_i \leq X_{i+2} \leq \ldots \leq X_k < b\}.$$

Then

(4.4.5) $$B_i(x) = \int_x^\infty \int_u^b \int_v^b \ldots \int_{x_{k-1}}^b f_i(u)f_{i+1}(v)\ldots f_k(x_k)dudv \ldots dx_k$$

$$\geq \int_x^\infty \int_u^b \int_v^b \ldots \int_{x_{k-1}}^b f_{i+1}(u)f_i(v)\ldots f_k(x_k)dudv \ldots dx_k$$

$$= B_i^*(x).$$

On the other hand,

$$P_r\{a < X_1 \leq \ldots \leq X_i \leq X_{i+1} \leq \ldots \leq X_k < b\}$$

$$= E_{\theta_{i-1}}\{A_i(X_{i-1})B_i(X_{i-1})\}$$

$$\geq E_{\theta_{i-1}}\{A_i(X_{i-1})B_i^*(X_{i-1})\}$$

$$= P_r\{a < X_1 \leq \ldots \leq X_{i+1} \leq X_i \leq \ldots \leq X_k < b\}.$$

This proves the lemma.

Remark 4.4.1. It should be pointed out that the above result will no longer be true if we weaken the condition by replacing condition "having monotone likelihood ratio property" by "which is stochastically increasing". For example, let X_1 and X_2 be two independently distributed random variables with mass function defined as follows:

	0	1	2	3
X_1	$\frac{1}{2}$	0	0	$\frac{1}{2}$
X_2	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Clearly X_2 is stochastically larger than X_1. But

$$P_r\{1 < X_1 \leq X_2 < 4\} = \frac{1}{4} < \frac{3}{8} = P_r\{1 < X_2 \leq X_1 < 4\}.$$

Theorem 4.4.3. If $X_1,\ldots,X_k$ are k independently distributed random variables with parameter $\theta_1,\ldots,\theta_k$ such that $\theta_1 \leq\ldots\leq \theta_k$, respectively, and with densities having monotone likelihood ratio property, and if a and b are two (extended) real numbers such that a < b, then

$$(4.4.6)\quad P_r\{a < X_k \leq\ldots\leq X_1 < b\} \leq P_r\{a < X_{\alpha_1} \leq\ldots\leq X_{\alpha_k} < b\} \leq P_r\{a < X_1 \leq\ldots\leq X_k < b\}$$

where $(\alpha_1,\ldots,\alpha_k)$ is any permutation of $(1,\ldots,k)$.

Proof. Define a mapping among the ordered k-tuples as follows: Let $T(i,j)$ be the mapping that interchanges the positions of X_i and X_j; that is, $T(\alpha_i,\alpha_j)\{X_{\alpha_1} \leq\ldots\leq X_{\alpha_i} \leq\ldots\leq X_{\alpha_j} \leq\ldots\leq X_{\alpha_k}\} = \{X_{\alpha_1} \leq\ldots\leq X_{\alpha_j} \leq\ldots\leq X_{\alpha_i} \leq\ldots\leq X_{\alpha_k}\}$. It is easy to see that any ordered k-tuple $\{X_{\alpha_1} \leq\ldots\leq X_{\alpha_k}\}$ can be obtained by applying finitely many times mappings of the form $T(\beta_i,\beta_{i+1})$ with $\beta_i < \beta_{i+1}$ on the ordered k-tuples $\{X_1 \leq\ldots\leq X_k\}$. By Lemma 4.4.2, whenever we apply $T(\beta_i, \beta_{i+1})$ on $\{X_{\beta_1} \leq\ldots\leq X_{\beta_k}\}$, the probability $P_r\{a < X_{\beta_1} \leq\ldots\leq X_{\beta_i} \leq X_{\beta_{i+1}} \leq\ldots\leq X_{\beta k} < b\}$ will decrease to

$P_r\{a < X_{\beta_1} \leq \ldots \leq X_{\beta_{i+1}} \leq X_{\beta_i} \leq \ldots \leq X_{\beta_k} < b\}$. This proves the theorem.

Corollary 4.4.1. Let a and b be two (extended) real numbers such that $a < b$, and let $X_1,\ldots,X_k$ be k independent binomial random variables with parameter $(n,p_1),\ldots,(np_k)$ respectively. Write $\theta_i = \varphi(p_i)$, $i = 1,\ldots,k$. If $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$, then

$$(4.4.7)\quad P_r\{a < \varphi(\frac{X_k}{n}) \leq \ldots \leq \varphi(\frac{X_1}{n}) < b\} \leq P\{a < \varphi(\frac{X_{\alpha_1}}{n}) \leq \ldots \leq \varphi(\frac{X_{\alpha_k}}{n}) < b\} \leq P_r\{a < \varphi(\frac{X_1}{n}) \leq \ldots \leq \varphi(\frac{X_k}{n}) < b\}$$

where $(\alpha_1,\ldots,\alpha_k)$ is any permutation of $(1,\ldots,k)$.

We state without proof other properties of the selection procedure R_φ.

Lemma 4.4.3. Let $\Theta_0 = \{\underline{\theta} = (\theta,\ldots,\theta),\ 0 \leq \theta \leq \varphi(\frac{1}{2})\}$. For any $\theta \in \Theta_0$ and a given constant $d > 0$,

$$(4.4.8)\qquad \lim_{n\to\infty} P_{\underline{\theta}}(CS|R_\varphi) = 1$$

Lemma 4.4.4. Let $\delta > 0$ and let $\Theta(\delta) = \{\underline{\theta} = (\theta_1,\ldots,\theta_k),\ \theta_{[k]} \geq \theta_{[k-1]} + \delta\}$. If $0 < d < \frac{1}{2}\delta$, then for any $\underline{\theta} \in \Theta(\delta)$,

$$(4.4.9)\qquad \lim_{n\to\infty} P_{\underline{\theta}}(ICD|R_\varphi) = 0$$

where ICD means incorrect decision, i.e., to include any non best population in the selected subset.

Corollary 4.4.2. For any $\underline{\theta} \in \Theta(\delta)$ and $0 < d < \frac{1}{2}\delta$,

(4.4.10) $$\lim_{n\to\infty} E_{\underline{\theta}}(S|R_\varphi) = 1.$$

Where S is the total number of population that are included in the selected subset.

4.5 EXPECTED SUBSET SIZE OF THE PROCEDURE

Consistent with the basic probability requirement, we would like the size of the selected subset to be small. Now S, the size of the selected subset is a random variable which takes values 1,2,...,k. Hence one criterion of the efficiency of the procedure R_φ is the expected value of the size of the subset.

Let Δ be a positive number such that $0 < \Delta < \varphi(\frac{1}{m},\ldots,\frac{1}{m})$, and let $\underline{p} = (p_1,\ldots,p_m)$ and $\underline{q} = (q_1,\ldots,q_m)$ be two vectors such that $\varphi(p_1,\ldots,p_m) = \Delta$ and $\varphi(q_1,\ldots,q_m) = \varphi(\frac{1}{m},\ldots,\frac{1}{m}) - \Delta$. Let $\Omega_\Delta = \{\underline{\omega} = (\underline{p}_1,\ldots,\underline{p}_k),\ \underline{p} \prec p_i \prec \underline{q}\ \ \forall_i\}$. Then we have the following.

Theorem 4.5.1.

(4.5.1) $$\sup_{\Omega_\Delta} E(S|R_\varphi) \le k\{\Sigma^*\binom{n}{x_1,\ldots,x_m}\prod_{i=1}^{m} p_i^{x_i}[\Sigma^{**}\binom{n}{y_1(\underline{x}),\ldots,y_m(\underline{x})} \cdot \prod_{i=1}^{m} \underline{q}_i^{y_i(\underline{x})}]^{k-1}\}$$

where the sum Σ^* extends over all m-tuples $\underline{x} = (x_1,\ldots,x_m)$ of non-negative integers for which $\sum_{i=1}^{m} x_i = n$ and the sum Σ^{**} extends over $(y_1,\ldots,y_m)$ as in Σ^* satisfying the additional condition

$$\varphi(\frac{y_1}{n},\ldots,\frac{y_m}{n}) \le \varphi(\frac{x_1}{n},\ldots,\frac{x_m}{n}) + d.$$

Proof. For any $\underline{\omega} \in \Omega_\Delta$,

$$(4.5.2)\quad E_{\underline{\omega}}(S|R_\varphi)$$

$$= \sum_{i=1}^{k} P_{\underline{\omega}}(\varphi(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}) \ge \max_{1\le j\le k} \varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}) - d)$$

$$= \sum_{i=1}^{k} \Sigma^*\binom{n}{x_{i1},\ldots,x_{im}} \prod_{\ell=1}^{m} p_{i\ell}^{x_{i\ell}} \{ \prod_{\substack{1\le j\le k \\ j\ne i}} P_{\underline{\omega}}(\varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}) \le \varphi(\frac{x_{i1}}{n},\ldots,\frac{x_{im}}{n})+d)\}$$

since

$$(4.5.3)\quad P_{\underline{\omega}}(\varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{j1}}{n}) \le \varphi(\frac{x_{i1}}{n},\ldots,\frac{x_{im}}{n})+d) \le \Sigma^{**}\binom{n}{y_1,\ldots,y_m} \prod_{i=1}^{m} q_i^{y_i}.$$

Also $P_{\underline{\omega}}(\varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}) \le \varphi(\frac{x_{i1}}{n},\ldots,\frac{x_{im}}{n})+d)$ is a Schur-concave function in $(x_{i1},\ldots,x_{im})$. The result follows.

4.6 SELECTING A SUBSET WHICH CONTAINS ALL POPULATIONS BETTER THAN A CONTROL

In this section, we discuss a related problem of selection using the same notation as described in Section 4.2, let

$$\pi_i = \begin{pmatrix} A_1 & \cdots & A_m \\ \\ p_{i1} & \cdots & p_{im} \end{pmatrix}, \qquad i = 0,1,\ldots,k,$$

be k+1 independent finite schemes. The procedure described in this section control the probability that the subset contains all those populations better than the standard ($\underline{p}_i \prec \underline{p}_0$), with the probability of a correct decision to be at least P*. Let X_{ij} denote the number of outcomes of $j\underline{th}$ event in scheme π_i, $j = 1,\ldots,m$, $i = 0,1,\ldots,k$. Let $(X'_{i1},\ldots,X'_{im})$ denote the events associated with the scheme $\pi_{(i)}$ with $\underline{p}'_i$ for which $\underline{p}' \prec \underline{p}_0$, $i = 1,\ldots,r_1$ where r_1 represents the number of schemes better than scheme π_0. We discuss the following cases.

CASE (A). KNOWN CONTROL

We assume $\underline{p}_0$ is known, and propose a class of procedures as follows:

R'_φ: Select scheme π_i if and only if

$$\varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) \geq \varphi(p_{01},\ldots,p_{0,1}) - c_1. \tag{4.6.1}$$

Let P_1 denote the probability of a correct decision. Then

$$\begin{aligned} P_1 &= \prod_{i=1}^{r_1} P_r\left(\varphi\left(\frac{X'_{i1}}{n},\ldots,\frac{X'_{im}}{n}\right) \geq \varphi(p_{01},\ldots,p_{0m}) - c_1\right) \\ &\geq \left\{P_r\left(\varphi\left(\frac{X_{01}}{n},\ldots,\frac{X_{0m}}{n}\right) \geq \varphi(p_{01},\ldots,p_{0m}) - c_1\right)\right\}^k. \\ &= \left\{\Sigma^*\binom{n}{x_1,\ldots,x_m} \prod_{i=1}^{m} p_{0_i}^{x_i}\right\}^k \end{aligned} \tag{4.6.2}$$

where the sum Σ^* is over all m-tuples $(x_1,\ldots,x_m)$ of nonnegative integers such that $\sum_{i=1}^{m} x_i = n$ and $\varphi\left(\frac{x_1}{n},\ldots,\frac{x_m}{n}\right) \geq \varphi(p_{01},\ldots,p_{0m}) - c_1$.

CASE (B) UNKNOWN CONTROL

Assume that $(p_{01},\ldots,p_{0m})$ is unknown. We propose in this case the following procedure:

R''_φ: Retain in the selected subset those and only those scheme π_i for which

$$\varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) \geq \varphi\left(\frac{X_{01}}{n},\ldots,\frac{X_{0m}}{n}\right) - c_2 \tag{4.6.3}$$

where c_2 is a constant such that $0 \leq c_2 \leq \varphi(\frac{n}{m},\ldots,\frac{n}{m})$.

The probability P_2 of a correct decision is given by

$$P_2 = \prod_{i=1}^{r_1} P_r\left(\varphi\left(\frac{X'_{i1}}{n},\ldots,\frac{X'_{im}}{n}\right) \geq \varphi\left(\frac{X_{01}}{n},\ldots,\frac{X_{0m}}{n}\right) - c_2\right). \tag{4.6.4}$$

$$\geq \left\{P_r\left(\varphi\left(\frac{Y_1}{n},\ldots,\frac{Y_m}{n}\right) \geq \varphi\left(\frac{X_{01}}{n},\ldots,\frac{X_{0m}}{n}\right) - c_2\right)\right\}^k$$

where $(Y_1,\ldots,Y_m)$ is a multinomial random vector with parameter n and $\underline{p}_0$ and $(Y_1,\ldots,Y_m)$ and $(X_{01},\ldots,X_{0m})$ are stochastically independent.

For a given fixed vector $\underline{t} = (t_1,\ldots,t_m)$, let $c_2(\underline{t})$ be the smallest number such that

$$M(2,c_2(\underline{t}),\underline{t},m,n) \geq \binom{2n}{t_1,\ldots,t_m} \sqrt[k]{P^*} \tag{4.6.5}$$

Let $c_2 = \max\{c_2(\underline{t})\colon\ t_i$ is a nonnegative integer, $i = 1,\ldots,k$ and $\sum_{i=1}^{m} t_i = n\}$,

we have $P_2 \geq P^*$.

4.7 APPLICATION

Let $\pi_i = \begin{pmatrix} A_1 & \cdots & A_m \\ p_{i1} & \cdots & p_{im} \end{pmatrix}$, $i = 1,2,\ldots,k$

be k independent finite schemes. In some practical situations, one wishes to know whether these finite schemes are significantly different or not. That is to decide whether the components of these k finite schemes differ only by permutations. Or equivalently, one wishes to test the hypothesis H_0: the set of vectors $\{(p_{i1},\ldots,p_{im}),\ i = 1,\ldots,k\}$ differ only by permutations.

Suppose we are allowed to take n independent observations from each of the k schemes. Let $(X_{i1},\ldots,X_{im})$ be the data obtained from the i<u>th</u> scheme, where X_{ij} represents the total number of outcomes of the j<u>th</u> event in the scheme π_i. Intuitively, we reject the hypothesis whenever there are significant differences among the values of $\varphi(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n})$, $i = 1,\ldots,k$. In other words, we reject the hypothesis if

$$(4.7.1) \qquad \max_{1\leq i\leq k} \varphi(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}) - \min_{1\leq i\leq k} \varphi(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}) > c$$

where c is chosen such that the probability of rejecting the hypothesis under H_0 is no more than a preassigned value α. A conservative constant c can be obtained as follows.

For a given α, $0 < \alpha < 1$, let c be the selection constant such that

$$(4.7.1) \qquad \inf_{\Omega} P(CS|R_\varphi) \geq 1 - \frac{\alpha}{k} .$$

Then under H_0,

$$P_r\left(\max_{1\le i\le k} \varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) - \min_{1\le i\le k} \varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) > c\right)$$

$$= P_r\left(\varphi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right) < \max_{1\le i\le k} \varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) - c, \text{ for some } j,\ 1\le j\le k\right)$$

$$\le k\, P_r\left(\varphi\left(\frac{X_{k1}}{n},\ldots,\frac{X_{km}}{n}\right) < \max_{1\le i\le k-1} \varphi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) - c\right)$$

$$\le k\{1-P_{\Omega_0}(CS|R_\varphi)\}$$

$$\le k(1-(1-\tfrac{\alpha}{k}))$$

$$= \alpha.$$

4.8 SELECTION PROCEDURES FOR THE LEAST UNCERTAINTY OF SEVERAL FINITE SCHEMES

Suppose the k independent finite schemes are such that one of them has the least uncertainty in the sense that the corresponding parameter vector majorizes the other parameter vectors. This implies that the associated measure of uncertainty H has the smallest value of $H_1,\ldots,H_k$. Suppose that we take n independent observations from each of the schemes. Let $(X_{i1},\ldots,X_{im})$ be the data obtained from scheme π_i, and let ψ be a Schur-convex function. We propose a selection procedure as follows:

R_ψ: Select scheme π_i if and only if

$$(4.8.1)\quad \psi\left(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}\right) \ge \max_{1\le j\le k} \psi\left(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}\right)-d'.$$

Note that if ψ is Schur-convex then $-\psi$ is Schur-concave. Hence by letting $\varphi = -\psi$, the above procedure can be rewritten as follows: Select scheme π_i if and only if

$$(4.8.2) \qquad \varphi(\frac{X_{i1}}{n},\ldots,\frac{X_{im}}{n}) \le \min_{1\le j\le k} \varphi(\frac{X_{j1}}{n},\ldots,\frac{X_{jm}}{n}) + d'.$$

By using similar arguments as in the previous section, it can be shown that the infimum of the probability of a correct selection occurs when the k given schemes are identical. Moreover results parallel to R_φ follows with obvious modifications.

TABLE I

Table of values d needed to satisfy the P* condition of the selection procedure R_φ for k = 2(1)5, n = 2(1)10, and $\varphi(x) = x(1-x)$. The top number is the smallest value of d satisfying the requirements of Theorem 4.3.2 for P* = 0.75, the second for P* = 0.80, the third for P* = 0.90, and the bottom for P* = 0.95.

k \ n	2	3	4	5	6	7	8	9	10
2	0.250	0.222	0.188	0.160	0.139	0.123	0.125	0.124	0.120
	0.250	0.222	0.250	0.240	0.222	0.204	0.125	0.124	0.160
	0.250	0.222	0.250	0.240	0.222	0.204	0.140	0.222	0.210
	0.250	0.222	0.250	0.240	0.250	0.245	0.234	0.222	0.210
3	0.250	0.222	0.188	0.240	0.222	0.204	0.188	0.173	0.160
	0.250	0.222	0.250	0.240	0.222	0.204	0.188	0.173	0.160
	0.250	0.222	0.250	0.240	0.222	0.204	0.234	0.222	0.210
	0.250	0.222	0.250	0.240	0.250	0.245	0.234	0.222	0.240
4	0.250	0.222	0.250	0.240	0.222	0.204	0.188	0.173	0.160
	0.250	0.222	0.250	0.240	0.222	0.204	0.188	0.173	0.160
	0.250	0.222	0.250	0.240	0.222	0.245	0.234	0.222	0.210
	0.250	0.222	0.250	0.240	0.250	0.245	0.234	0.222	0.240
5	0.250	0.222	0.250	0.240	0.222	0.204	0.188	0.173	0.160
	0.250	0.222	0.250	0.240	0.222	0.204	0.188	0.173	0.160
	0.250	0.222	0.250	0.240	0.250	0.245	0.234	0.222	0.210
	0.250	0.222	0.250	0.240	0.250	0.245	0.234	0.222	0.240

TABLE II

Table of values of d needed to satisfy the P*-condition of the selection procedure R_φ for k = 2(1)5, n = 2(1)10, and $\varphi(x)=\min\{x,1-x\}$. The top number is the smallest value of d satisfying the requirements of Theorem 4.3.2 for P* = 0.75, the second for P* = 0.80, the third for P* = 0.90, and the bottom for P* = 0.95.

k \ n	2	3	4	5	6	7	8	9	10
2	0.500	0.333	0.250	0.200	0.333	0.286	0.250	0.222	0.200
	0.500	0.333	0.500	0.400	0.333	0.286	0.250	0.222	0.200
	0.500	0.333	0.500	0.400	0.333	0.286	0.375	0.333	0.300
	0.500	0.333	0.500	0.400	0.500	0.429	0.375	0.333	0.400
3	0.500	0.333	0.250	0.200	0.333	0.286	0.250	0.222	0.200
	0.500	0.333	0.500	0.400	0.333	0.286	0.250	0.222	0.200
	0.500	0.333	0.500	0.400	0.333	0.286	0.375	0.333	0.300
	0.500	0.333	0.500	0.400	0.500	0.429	0.375	0.333	0.400
4	0.500	0.333	0.500	0.400	0.333	0.286	0.250	0.222	0.200
	0.500	0.333	0.500	0.400	0.333	0.286	0.250	0.222	0.300
	0.500	0.333	0.500	0.400	0.333	0.429	0.375	0.333	0.300
	0.500	0.333	0.500	0.400	0.500	0.429	0.375	0.333	0.400
5	0.500	0.333	0.500	0.400	0.333	0.286	0.250	0.222	0.200
	0.500	0.333	0.500	0.400	0.333	0.286	0.250	0.222	0.333
	0.500	0.333	0.500	0.400	0.333	0.429	0.375	0.333	0.300
	0.500	0.333	0.500	0.400	0.333	0.429	0.375	0.333	0.400

CHAPTER V

ON SELECTION AND RANKING PROCEDURES FROM THE NEGATIVE MULTINOMIAL DISTRIBUTION AND SOME RELATED PROBLEMS

5.1 INTRODUCTION

We consider a sequence of independent trials, in each of which the event A_i occurs with probability p_i $(i = 0,1,\ldots,k;$ $\sum_{i=0}^{k} p_i = 1)$. Let X_i be the frequency of A_i before the rth appearance of A_0. Then $(X_1,\ldots,X_k)$ has the negative multinomial distribution

$$(5.1.1)\quad P(X_1 = x_1,\ldots,X_k = x_k) = \frac{\Gamma(r + \sum_{i=1}^{k} x_i)}{\Gamma(r)} p_0^r \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!} .$$

This model has been proposed and used in the statistical theory of accident proneness, absenteeism and contagion.

Recently some selection procedures have been proposed and studied by Gupta and Huang [39], Gupta and Nagel [44] and Panchapakesan [70] for the multinomial distribution. However, very little work has been done on the selection problem for the negative multinomial distribution. In this chapter, some selection procedures for the negative multinomial are proposed and studied. Let $p_{[1]} \leq \cdots \leq p_{[k]}$ denote the ordered values of the p_i. Given any P^*, $\frac{1}{k} < P^* < 1$, we want to select a subset of these k events such that the subset contains the one corresponding to the parameter $p_{[1]}$ with probability at least P*, no matter what the configuration of $p_1,\dots,p_k$ is. We denote this by CS. Therefore we are interested in defining a selection rule R such that

$$\inf_{\Omega} P(CS|R) \geq P^*, \tag{5.1.2}$$

where $\Omega = \{(p_1,\dots,p_k):\ p_i \geq 0,\ 0 < \sum_{i=1}^{k} p_i < 1\}$. Here we assume that the parameter p_0 and r are known. In Section 5.2 two unconditional rules are proposed and studied. Conditional procedures are considered in Section 5.3. An inverse-sampling procedure for selecting a cell in a multinomial distribution is studied in Section 5.4.

5.2. UNCONDITIONAL SELECTION PROCEDURES R_{10} AND R_{11}

To select a subset that contains the cell with lowest p_i, we propose two unconditional procedures. We observe the vector $\underline{X} = (X_1,\dots,X_k)$. First of all, we define a rule as follows.

R_{10}: Select π_i if and only if

$$X_i \leq c \min_{1\leq j\leq k} X_j + (c-1)r \tag{5.2.1}$$

where $c \geq 1$ is the smallest number such that the basic probability requirement is satisfied.

Before we investigate the least favorable configuration of P(CS), we derive first the following useful lemma.

Lemma 5.2.1. Let $\underset{\sim}{X} = (X_1,\ldots,X_k)$ have the negative multinomial distribution

$$(5.2.2)\quad P(x_1,\ldots,x_k) = \frac{\Gamma(r + \sum_{i=1}^{k} x_i)}{\Gamma(r)}\, p_0^r \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!}$$

with $0 < p_0 < 1$, and let $\phi(x_1,\ldots,x_k)$ be a symmetric Schur-concave function such that $E\phi(X_1,\ldots,X_k)$ exists, then $E\phi(X_1,\ldots,X_k)$ is Schur-concave in $(p_1,\ldots,p_k)$.

Proof. Let $\psi(p_1,\ldots,p_k) = E\phi(X_1,\ldots,X_k)$

$$= \sum_{x_1=0}^{\infty}\cdots\sum_{x_k=0}^{\infty} \frac{\Gamma(r+x_1+\ldots+x_k)}{\Gamma(r)}\, p_0^r\phi(x_1,\ldots,x_k) \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!} .$$

Suppose $p_1 \geq p_2$,

$$\frac{\partial\psi}{\partial p_1} - \frac{\partial\psi}{\partial p_2} = \sum_{x_1=0}^{\infty}\cdots\sum_{x_k=0}^{\infty} \frac{\Gamma(r+x_1+\ldots+x_k)}{\Gamma(r)}\, p_0^r\phi(x_1,\ldots,x_k)\left(\frac{x_1}{p_1} - \frac{x_2}{p_2}\right) \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!}$$

$$= \sum_{x_1=0}^{\infty}\sum_{x_k=0}^{\infty} \frac{\Gamma(r+x_1+\ldots+x_k)}{\Gamma(r)}\, p_0^r\{\phi(x_1+1,x_2,\ldots,x_k) - \phi(x_1,x_2+1,\ldots,x_k)\} \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!} .$$

Since $\phi(x_1+1,x_2,\ldots,x_k) = \phi(x_1,x_2+1,\ldots,x_k)$ if $x_1 = x_2$, we can split the above summation into three parts, namely sum over the cases when $x_1 > x_2$, $x_1 < x_2$ and $x_1 = x_2$. Then we have

$$\frac{\partial\psi}{\partial p_1} - \frac{\partial\psi}{\partial p_2} = p_0^r \sum_{x_1>x_2}\cdots\sum \frac{\Gamma(r+1+x_1+\ldots+x_k)}{\Gamma(r)} \{\phi(x_1+1,x_2,\ldots,x_k) - \phi(x_1,x_2+1,\ldots,x_k)\} \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!}$$

$$+ p_0^r \sum_{x_1<x_2}\cdots\sum \frac{\Gamma(r+1+x_1+\ldots+x_k)}{\Gamma(r)} \{\phi(x_1+1,x_2,\ldots,x_k) - \phi(x_1,x_2+1,\ldots,x_k)\} \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!}$$

$$= p_0^r \sum_{x_1>x_2}\cdots\sum \frac{\Gamma(r+1+x_1+\ldots+x_k)}{\Gamma(r)\prod_{i=1}^{k} x_i!} (p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k})\{\phi(x_1+1,x_2,\ldots,x_k) - \phi(x_1,x_2+1,\ldots,x_k)\}$$

$$+ p_0^r \sum_{x_1>x_2}\cdots\sum \frac{\Gamma(r+1+x_1+\ldots+x_k)}{\Gamma(r)\prod_{i=1}^{k} x_i!} (p_1^{x_2} p_2^{x_1} \ldots p_k^{x_k})\{\phi(x_2+1,x_1,\ldots,x_k) - \phi(x_2,x_1+1,\ldots,x_k)\}$$

$$= p_0 \sum_{x_1>x_2}\cdots\sum \frac{\Gamma(r+1+x_1+\ldots+x_k)}{\Gamma(r)\prod_{i=1}^{p} x_i!} (p_1^{x_1} p_2^{x_2} p_3^{x_3} \ldots p_k^{x_k} - p_1^{x_2} p_2^{x_1} p_3^{x_3} \ldots p_k^{x_k}) (\phi(x_1+1,x_2,\ldots,x_k) - \phi(x_1,x_2+1,\ldots,x_k)).$$

Since $(x_2+1,x_1,x_3,\ldots,x_k) \prec (x_1+1,x_2,x_3,\ldots,x_k)$ and $p_1^{x_1} p_2^{x_2} - p_1^{x_2} p_1^{x_1} = (p_1 p_2)^{x_2}(p_1^{x_1-x_2} - p_2^{x_1-x_2}) \geq 0$ if $x_1 \geq x_2$. It follows that $\frac{\partial\psi}{\partial p_1} - \frac{\partial\psi}{\partial p_2} \geq 0$. Hence this proves the lemma.

Now let $X_{(i)}$ be the unknown statistic associated with $p_{[i]}$, $i = 1,\ldots,k$. For any configuration $\underline{p} = (p_1,\ldots,p_k)$, let

(5.2.3) $P_{\underline{p}}(i) = P_{\underline{p}}(\pi_{(i)} \text{ is selected} | R)$.

Let $(x_1,\ldots,\hat{x}_i,\ldots,x_k)$ denote the (k-1)-tuple obtained by deleting the $i\underline{th}$ component from $(x_1,\ldots,x_i,\ldots,x_k)$, and let

$$(5.2.4)\quad h_{x_i}(x_1,\ldots,\hat{x}_i,\ldots,x_k) = \begin{cases} 1 & \text{if } c \min\limits_{\substack{1\le j\le k\\ j\ne i}} x_j \ge x_i-(c-1)r \\ 0 & \text{otherwise.} \end{cases}$$

Then h_{x_i} is a symmetric Schur-concave function of $(x_1,\ldots,\hat{x}_i,\ldots,x_k)$. Furthermore, given $X_{(i)} = x_i$, the (k-1)-dimensional random vector $(X_{(1)},\ldots,\hat{X}_{(i)},\ldots,X_{(k)})$ is distributed as negative multinomial with parameters $r+x_i$ and $(p_{[1]},\ldots,\hat{p}_{[i]},\ldots,p_{[k]})$. By Lemma 5.2.1, given $X_{(i)} = x_i$, $Eh_{x_i}(X_{(1)},\ldots,\hat{X}_{(i)},\ldots,X_{(k)})$ is Schur-concave in $(p_{[1]},\ldots,\hat{p}_{[i]},\ldots,p_{[k]})$. Since $P(CS|R_{10}) = P_{\underline{p}}(1)$, and $(p_{[1]},p_{[1]},\ldots,p_{[1]},q) \succ (p_{[2]},\ldots,p_{[k]})$ whenever $(k-2)p_{[1]} + q = \sum_{i=2}^{k} p_{[i]}$, and $p_{[1]} \le q$. Hence we have shown the first part of the following theorem.

<u>Theorem 5.2.1</u>. The least favorable configuration of the parameters for the rule R_1 is of the form $(p,\ldots,p,q)$, where $p \le q$, $(k-1)p + q = 1-p_0$. Furthermore, if $k = 2$, then $p = q = \frac{1}{2}(1-p_0)$.

<u>Proof</u>. To complete the proof, consider the case $k = 2$, then

$$(5.2.5)\quad P(CS|R_{10}) = \sum_{t=0}^{\infty} \frac{\Gamma(r+t)}{\Gamma(r)t!} p_0^r(1-p_0)^t \left\{ \sum_{i=0}^{\left[\frac{c(t+r)-r}{1+c}\right]} \binom{t}{i} \left(\frac{p_{[1]}}{p_{[1]}+p_{[2]}}\right)^i \right\} \left(\frac{p_{[2]}}{p_{[1]}+p_{[2]}}\right)^{t-i} .$$

It follows that the infimum of $P(CS|R_1)$ takes place when $p_{[1]} = p_{[2]} = \frac{1}{2}(1-p_0)$.

By using the fact the conditional distribution of $(X_1,\ldots,X_k)$ given $\sum_{i=1}^{k} X_i = t$ is a multinomial distribution with parameters t and $(\frac{p_1}{1-p_0},\ldots,\frac{p_k}{1-p_0})$, we conclude that R_1 is a reverse monotone procedure. Next let us show how to obtain a conservative value of the selection constant.

<u>Theorem 5.2.2</u>. For a given P* and any $t \geq 0$, let $P_1^* = 1 - \frac{1-P^*}{k-1}$, let c(t) be the largest value such that $I_{\frac{1}{2}}(\frac{t-cr+r}{1+c}, \frac{c(1+t)+r(c-1)}{1+c}) \geq P_1^*$, where $I_p(a,b)$ is defined to be 1 if $a < 0$. If $c = \inf_{t\geq 0}\{c(t)\}$, then $\inf_{\Omega} P(CS|R_{10}) \geq P^*$.

<u>Proof</u>.

$$
\begin{aligned}
P(CS|R_{10}) &= P(X_{(1)} \leq c \min_{2\leq j\leq k} X_{(j)}+(c-1)r) \\
&= 1-P(X_{(1)} > cX_{(j)}+(c-1)r \quad \text{for some } 2 \leq j \leq k) \\
&\geq 1-\sum_{j=2}^{k}(1-P(X_{(1)} \leq cX_{(j)} + (c-1)r)) \\
&= 2-k + \sum_{j=2}^{k}\sum_{t=0}^{\infty} P(X_{(j)} \geq \tfrac{t-cr+r}{1+c}|X_{(1)} + X_{(j)} = t)P(X_{(1)}+X_{(j)}=t) \\
&\geq 2-k + (k-1)P_1^* \\
&= P^*
\end{aligned}
$$

Thus, the proof is complete.

Let Ω_1 be the set of k-tuples in the parameter space for which $p_{[1]} \geq \delta > 0$, where $0 < \delta < \frac{1}{k}(1-p_0)$. We discuss the expected subset size as follows:

Theorem 5.2.3.

$$(5.2.6)\quad \sup_{\Omega_1} E(S|R_{10}) \leq k\{(\frac{p_0}{p_0+2\delta})^r + \sup_{t\geq 1} g(t)(1-(\frac{2\delta}{p_0+2\delta})^r)\}$$

where $g(t) = \sum_{i=0}^{[\frac{ct+(c-1)r}{1+c}]} \binom{t}{i}(\frac{\delta}{1-p_0})^i(\frac{1-p_0-\delta}{1-p_0})^{t-i}$.

Proof.

$$E_{\Omega_1}(S|R_{10}) = \sum_{i=1}^{k} P(X_{(i)} \leq c \min_{1\leq j\leq k} X_{(j)}+(c-1)r)$$

$$\leq \frac{1}{k-1}\sum_{i=1}^{k}\sum_{j\neq i} P(X_{(i)} \leq cX_{(j)}+(c-1)r)$$

$$= \frac{1}{k-1}\sum_{i=1}^{k}\sum_{j\neq i}\sum_{t=0}^{\infty} P(X_{(i)} \leq \frac{ct+(c-1)r}{1+c}|X_{(i)}+X_{(j)}=t)P(X_{(i)} + X_{(j)} = t)$$

$$= \frac{1}{k-1}\sum_{i=1}^{k}\sum_{j\neq i}\sum_{t=0}^{\infty} g(t)\frac{\Gamma(r+t)}{\Gamma(r)t!}(\frac{p_0}{p_0+p_{[i]}+p_{[j]}})^r(\frac{p_{[i]}+p_{[j]}}{p_0+p_{[i]}+p_{[j]}})^t$$

$$\leq \frac{1}{k-1}\sum_{j=1}^{k}\sum_{j\neq i}\{(\frac{p_0}{p_0+2\delta})^r + \sum_{t=1}^{\infty} g(t)\frac{\Gamma(r+t)}{\Gamma(r)t!}(\frac{p_0}{p_0+p_{[i]}+p_{[j]}})^r (\frac{p_{[i]}+p_{[j]}}{p_0+p_{[i]}+p_{[j]}})^t\}$$

$$\leq \frac{1}{k-1} \sum_{i=1}^{k} \sum_{j\neq i} \{(\frac{p_0}{p_0+2\delta})^r + \sup_{t\geq 1} g(t)(1-(\frac{2\delta}{p_0+2\delta})^r)\}$$

$$= k\{(\frac{p_0}{p_0+2\delta})^r + \sup_{t\geq 1} g(t)(1-(\frac{2\delta}{p_0+2\delta})^r)\}.$$

This completes the proof.

Next we propose another unconditional rule as follows

R_{11}: Select π_i if and only if

(5.2.7) $$X_i \leq \frac{c}{k-1} \sum_{j\neq i} x_j + (c-1)r.$$

In this case

$$P_\Omega(CS|R_{11}) = P_r(X_{(1)} \leq \frac{c}{k-1} \sum_{2}^{k} X_{(j)} + (c-1)r)$$

$$= \sum \frac{\Gamma(r+x+y)}{\Gamma(r)x!y!} p_0^r \, p_{[1]}^x (\sum_{j=2}^{k} p_{[j]})^y$$

where the summation is over all pair of non negative integers (x,y) such that $x \leq \frac{cy}{k-1} + (c-1)r$. By using the similar argument as in the proof of the second assertion of Theorem 5.2.1, we obtain the following theorem.

Theorem 5.2.4.

(5.2.8) $$\inf_{\Omega} P(CS|R_{11}) = \sum \frac{\Gamma(r+x+y)}{\Gamma(r)+x!y!} p_0^r (\frac{1-p_0}{k})^x (\frac{(k-1)(1-p_0)}{k})^y$$

where the summation is over all pair of nonnegative integers (x,y) such that $x \leq \frac{cy}{k-1} + (c-1)r$.

By using a similar argument as in the proof of Theorem 5.2.1, we have

(5.2.9) $P_{\underline{p}}(\pi_{(i)} \text{ is selected} | R_{11})$

$$= \sum_{t=0}^{\infty} \frac{\Gamma(r+t)}{\Gamma(r)t!} p_0^r (1-p_0)^t \{ \sum_{j=0}^{\left[\frac{ct+(k-1)(c-1)r}{k-1+c}\right]} \binom{t}{j} \left(\frac{p_{[i]}}{1-p_0}\right)^j \left(1 - \frac{p_{[i]}}{1-p_0}\right)^{t-j} \}.$$

It follows that

(5.2.10) $P_{\underline{p}}(\pi_{(i)} \text{ is selected} | R_{11}) \geq P_{\underline{p}}(\pi_{(j)} \text{ is selected} | R_{11})$

whenever $i \leq j$.

5.3 CONDITIONAL PROCEDURES R_{12} AND R_{13}

R_{12}: Select π_i if and only if

(5.3.1) $X_i \leq c \min_{1 \leq j \leq k} X_j + (c-1)r, \quad \text{given } \sum_1^k X_i = t.$

It has been pointed out that the conditional distribution of $(X_1,\ldots,X_k)$ given $\sum_1^k X_i = t$ is a multinomial distribution with parameter t and $(\frac{p_1}{1-p_0},\ldots,\frac{p_k}{1-p_0})$. By using an analogous argument as in the proof of Theorem 3, we conclude that the conditional procedure R_3 is reverse monotone and the least favorable configuration of $P(CS|R_{12})$ is given by

(5.3.2) $$p_{[\ell]} = \begin{cases} p, & \ell = 1,\ldots,k-1 \\ 1-(k-1)p, & \ell = k. \end{cases}$$

Theorem 5.3.1. For a given P*, ($\frac{1}{k}$ < P* < 1), let c($\geq$ 1) be the smallest number such that

$$\sum_{i=0}^{[\frac{t-(c-1)r}{1+c}]} \binom{t}{i} \frac{1}{2^t} \leq \frac{1-P^*}{k-1}.$$

Then $\inf_{\Omega} P(CS|R_{12}) \geq P^*$.

Proof. For any $\underset{\sim}{p} \in \Omega$,

$$P_{\underset{\sim}{p}}(CS|R_{12}) = P_r(X_{(1)} \leq c \min_{2\leq j\leq k} X_{(j)}+(c-1)r|\sum_{\ell=1}^{k} X_{(\ell)} = t)$$

$$= 1-P_r(X_{(1)} > c \min_{2\leq j\leq k} X_{(j)}+(c-1)r|\sum_{1}^{k} X_{(\ell)} = t)$$

$$\geq 1-\sum_{j=2}^{k} P_r(X_{(1)} > c\, X_{(j)}+(c-1)r|\sum_{1}^{k} X_{(\ell)} = t)$$

$$\geq 1-(k-1)P_r(X_{(1)} > cX_{(2)}+(c-1)r|\sum_{1}^{k} X_{(\ell)} = t)$$

$$\geq 1-(k-1) \sum_{i=0}^{[\frac{t-(c-1)r}{1+c}]} \binom{t}{i} \frac{1}{2^t}$$

$$\geq 1-(k-1)\cdot\frac{1-P^*}{k-1}$$

$$= P^*.$$

Let us denote by $\Omega(\delta)$, the space of all parameter vectors $(p_1,\ldots,p_k)$ such that $p_{[1]} \geq \delta > 0$. Then for any $\underset{\sim}{p} \in \Omega(\delta)$

$$E_{\underline{p}}(S|R_{12}) = \sum_{i=1}^{k} P_r(X_{(i)} \leq c \min_{1\leq j\leq k} X_{(j)}+(c-1)r| \sum_{i=1}^{k} X_{(i)}=t)$$

$$\leq \sum_{i=1}^{k} P_r(X_{(i)} \leq \frac{ct+(c-1)r}{1+c}| \sum_{i=1}^{k} X_{(i)} = t)$$

$$\leq k \sum_{i=0}^{[\frac{ct+(c-1)r}{1+c}]} \binom{t}{i}\delta^i(1-\delta)^{t-i}.$$

Hence

$$\sup_{\Omega(\delta)} E(S|R_{12}) \leq k \sum_{i=0}^{[\frac{ct+(c-1)r}{1+c}]} \binom{t}{i}\delta^i(1-\delta)^{t-i}.$$

Next we propose a conditional procedure as follows:

R_{13}: Select π_i if and only if

(5.3.3) $\quad X_i \leq \frac{c}{k-1} \sum_{j\neq i} X_j+(c-1)r, \quad \text{given} \quad \sum_{i=1}^{k} X_i = t.$

It is easy to see that (5.3.3) can be rewritten as follows:

(5.3.4) $\quad X_i \leq \frac{(k-1)(ct+(c-1)r)}{k-1+c}$.

Since the conditional distribution of X_i given $\sum_{i=1}^{k} X_i = t$ is a binomial distribution with parameters t and $\frac{p_i}{1-p_0}$. It follows that for any parameter vector $(p_1,\ldots,p_k)$,

(5.3.5) $\quad P(\pi_{(i)} \text{ is selected}|R_{13}) \geq P(\pi_{(j)} \text{ is selected}|R_{13})$

whenever $i < j$. And

$$\text{(5.3.6)} \quad \inf_{\Omega} P(CS|R_{13}) = \sum_{i=0}^{\left[\frac{(k-1)(ct+(c-1)r)}{k-1+c}\right]} \binom{t}{i}\left(\frac{1}{k}\right)^{i}\left(\frac{k-1}{k}\right)^{t-i}.$$

5.4 SOME RELATED SELECTION PROBLEM

The problem of selecting the particular one of the k multinomial cells with the highest probability was mentioned in Section 3.4, and a subset selection rule based on inverse sampling procedure in which the number of observations at each stage is a random number taken from a Poisson distribution. In this section the subset selection rules that use inverse sampling termination rules together with the vector-at-a-time sampling rule will be considered.

SELECTING THE MULTINOMIAL CELL WITH THE HIGHEST PROBABILITY

Observations are taken one at a time from the multinomial distribution until the count in any one of the cells reaches M. Let $X_1,\ldots,X_k$ be the cell counts at termination of sampling. Of course, one of the X_i is equal to M. Panchapakesan [70] has proposed the following rule:

R_{14}: Select the cell with the count X_i if and only if

$$\text{(5.4.1)} \quad X_i \geq M-D.$$

It is shown in [70] that the infimum of the probability of a correct selection $P(CS|R_{14})$ when rule R_{14} is used is attained for a configuration of the type $(0,\ldots,0,\frac{1}{r},\ldots,\frac{1}{r})$ where $r \geq 2$ is the number of non-zero cell-probabilities. An asymptotic result of $P(CS|R_{14})$ was also discussed in [70]. In this section, we are going to show that the least favorable configuration for the rule R_5 is $(\frac{1}{k},\ldots,\frac{1}{k})$. To show this, let X_i be the count associated with the cell π_i, and let $Y_j^{(i)}(M-D)$ denote the count of π_j before the (M-D)th appearance in π_i, $i,j = 1,\ldots,k$, $j \neq i$. If the cell-probabilities associated with π_i is θ_i, $i = 1,\ldots,k$, then $(Y_1^{(i)}(M-D),\ldots,Y_{i-1}^{(i)}(M-D),Y_{i+1}^{(i)}(M-D),\ldots,Y_k^{(i)}(M-D))$ has the negative multinomial distribution

$$(5.4.2)\quad P_r(Y_j^{(i)}(M-D) = y_j,\ j=1,\ldots,k,j\neq i) = \frac{\Gamma(M-D+\sum_{j\neq i} y_j)}{\Gamma(M-D)}\,\theta_i^{M-D}\prod_{j\neq i}\theta_j^{y_j}.$$

On the other hand, the rule R_{14} can be rewritten as follows: Select the cell π_i if and only if

$$(5.4.3)\qquad \max_{\substack{1\leq j\leq k\\ j\neq i}} Y_j^{(i)}(M-D) \leq M-1.$$

Let $\pi_{(i)}$ be the cell corresponding to $\theta_{[i]}$ and let $Y_{(j)}^{(i)}(M-D)$ denote the count of $\pi_{(j)}$ before the (M-D)th appearance in $\pi_{(i)}$. Then

$$P(CS|R_{14}) = P(\max_{1\leq j\leq k-1} Y_{(j)}^{(k)}(M-D) \leq M-1)$$

$$= \theta_{[k]}^{M-D}\sum_{y_1=0}^{M-1}\cdots\sum_{y_{k-1}=0}^{M-1}\frac{\Gamma(M-D+y_1+\ldots+y_{k-1})}{\Gamma(M-D)\prod_{i=1}^{k-1} y_i!}\prod_{i=1}^{k-1}\theta_{[i]}^{y_i}$$

$$(5.4.4) \quad = \frac{1}{B(M,\ldots,M,M-D)} \int_{\frac{\theta_{[1]}}{\theta_{[k]}}}^{\infty} \cdots \int_{\frac{\theta_{[k-1]}}{\theta_{[k]}}}^{\infty} \frac{(\prod_{i=1}^{k-1} y_i)^{M-1}}{(1+\sum_{i=1}^{k-1} y_i)^{k(M+1)-D-1}} \prod_{i=1}^{k-1} dy_i$$

(by Theorem 2.4 of [68])

where

$$B(a_1,\ldots,a_k) = \frac{\Gamma(a_1) \ldots \Gamma(a_k)}{\Gamma(a_1+\ldots+a_k)} .$$

It follows that the infimum of $P(CS|R_5)$ occurs at the configuration $(\frac{1}{k},\ldots,\frac{1}{k})$.

In other words,

$$(5.4.5) \quad \inf_{\Omega} P_{\underline{\theta}}(CS|R_{14}) = P(X_k \geq M-D|\theta_1=\ldots=\theta_k = \frac{1}{k})$$

where $\Omega = \{\underline{\theta} = (\theta_1,\ldots,\theta_k): 0 \leq \theta_i \leq 1,\ \theta_1 +\ldots+ \theta_k = 1\}$.

To investigate the properties of the rule R_5, let us introduce some definitions. Let $\underline{\theta} \in \Omega$, and define

$$(5.4.6) \quad P_{\underline{\theta}}(i) = P_{\underline{\theta}}(\pi_{(i)} \text{ is selected}|R).$$

From (5.4.3) and Lemma 5.2.1, it follows that $P_{\underline{\theta}}(i)$ is a Schur-concave function in $(\theta_{[1]},\ldots,\theta_{[i+1]},\theta_{[i+1]},\ldots,\theta_{[k]})$ when $\theta_{[i]}$ is fixed.

We say that the given vector $\underline{x} = (x_1,\ldots,x_n)$ weakly majorizes the vector $\underline{x}' = (x_1',\ldots,x_n')$ if

$$\sum_{i=1}^{j} x_{[n-i+1]} \geq \sum_{i=1}^{j} x'_{[n-i+1]}, \quad j = 1,\ldots,n;$$

in symbol, $\underline{x} \underline{>} \underline{>}^m \underline{x}'$. A random vector $\underline{X}$ is said to stochastically weakly majorize random vector $\underline{X}'$ if $f(\underline{X})$ is stochastically larger than $f(\underline{X}')$ for every nondecreasing Schur-convex function f; in symbol, $\underline{X} \underline{>} \underline{>}^{st,m,} \underline{X}'$. The following result is established by Nevius, Proschan and Sethuraman [67].

Lemma 5.4.1. The following statements are equivalent:

(i) $\underline{X} \underline{>} \underline{>}^{st,m,} \underline{X}'$

(ii) $Ef(\underline{X}) \geq Ef(\underline{X}')$ for every nondecreasing Schur-convex function f for which both these expectations exist.

Now let $(X_1,\ldots,X_n)$ have negative multinomial distribution

$$(5.4.7) \qquad P(X_1=x_1,\ldots,X_n=x_n) = \theta_0^r \frac{\Gamma(r+x_1+\ldots+x_n)}{\Gamma(r)} \prod_{i=1}^{n} \left(\frac{\theta_i^{x_i}}{x_i!}\right),$$

and let $\theta_0 = (1+\sum_{i=1}^{n} \lambda_i)^{-1}$, $\theta_j = \lambda_j(1+\sum_{i=1}^{n} \lambda_i)^{-1}$, $\lambda_j > 0$, $j = 1,\ldots,n$, then (5.4.7) can be rewritten as follows

$$(5.4.8) \quad P(X_1=x_1,\ldots,X_n=x_n) = \frac{\Gamma(r+\sum_{i=1}^{n} x_i)}{\Gamma(r)} (1+\sum_{i=1}^{n} \lambda_i)^{-r-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \left(\frac{\lambda_i^{x_i}}{x_i!}\right).$$

Nevius, Proschan and Sethuraman has also shown that if $\underline{X}_{\underline{\lambda}}$ has a negative multinomial distribution with density (5.4.8) and $\underline{\lambda} \underline{>} \underline{>}^m \underline{\lambda}'$, then $\underline{X}_{\underline{\lambda}} \underline{>} \underline{>}^{st,m,} X_{\underline{\lambda}}'$. Let us write $\theta_{[i]} = (1+\sum_{i=1}^{k-1} \lambda_i')^{-1}$, $\theta_{[j]} = \lambda_j(1+\sum_{i=1}^{k} \lambda_i')^{-1}$, $j = 1,\ldots,k$, $j \neq i$; and let $\theta_{[i+1]} = (1+\sum_{i=1}^{k-1} \lambda_i'')^{-1}$, $\theta_{[j]} = \lambda_j''(1+\sum_{i=1}^{k} \lambda_j'')^{-1}$, $j = 1,\ldots,k$, $j \neq i+1$. Then

$(Y_{(1)}^{(i)}(M-D),\ldots,Y_{(i-1)}^{(i)}(M-D),Y_{i+1}^{(i)}(M-D),\ldots,Y_{(k)}^{(i)}(M-D))$ has a negative multinomial distribution given by (5.4.8) with λ_i replaced by λ_i' and r by M-D. Similarly $(Y_{(1)}^{(i+1)}(M-D),\ldots,Y_{(i)}^{(i+1)}(M-D),Y_{(i+2)}^{(i+1)}(M-D),\ldots,Y_{(k)}^{(i+1)}(M-D))$ has a negative multinomial distribution with parameters M-D and $\underline{\lambda}'' = (\lambda_1'',\ldots,\lambda_k'')$. It is easily seen that $\underline{\lambda}'' \leq \leq^m \underline{\lambda}'$. Using the fact that $-I_{[\max(x_1,\ldots,x_{k-1}) \leq M-1]}$ is a nondecreasing Schur-convex function and Lemma 5.4.2, it follows that

$$(5.4.9) \qquad P_{\underline{\theta}}(i) \leq P_{\underline{\theta}}(i+1), \qquad i = 1,\ldots,k-1.$$

In other words, R_{14} is montone.

Next we discuss the expected subset size of R_5. For fixed M and D, let $\Omega_1 = \{\underline{\theta}: \theta_{[1]} \geq \frac{M-1-D}{k(M-1)-D}\}$. Suppose $\underline{\theta} \in \Omega_1$,

$$\begin{aligned} E_{\underline{\theta}}(S|R_{14}) &= \sum_{i=1}^{k} P_r(\max_{j\neq i} Y_j^{(i)}(M-D) \leq M-1) \\ &\leq \sum_{i=1}^{k} P_r(\sum_{j\neq i} Y_j^{(i)}(M-D) \leq k(M-1)) \\ &= \sum_{i=1}^{k} \theta_i^{M-D} \sum_{y=0}^{(k-1)(M-1)} \binom{M-D+y-1}{y}(1-\theta_i)^y \\ &= \sum_{i=1}^{k} I_{\theta_i}(M-D,(k-1)(M-1)+1) \\ &= Q(\underline{\theta}), \text{ say.} \end{aligned}$$

By using Ostrowski's theorem, it is easy to see that $Q(\underline{\theta})$ is a Schur-concave function in $\underline{\theta}$ when $\underline{\theta} \in \Omega_1$. This implies that

(5.4.10) $\sup_{\Omega_1} E_{\underline{\theta}}(S|R_{14}) \leq k\, I_{\frac{1}{k}}(M-D, (k-1)(M-D+1))$.

It should be pointed out that when k = 2, the upper bound given in (5.4.10) is an exact bound.

The average sample size E(N) for the procedure R_5 is given by

(5.4.11) $E(N) = M+\sum_{i=1}^{k} P(E_{(i)})\{\sum_{j\neq i} E(X_{(j)}|E_{(i)})\}$,

where $E_{(i)}$ is the event that the count in the cell $\pi_{(i)}$ reaches M first and $E(X_{(j)}|E_{(i)})$ is the conditional expectation of the count in cell $\pi_{(j)}$ given that $E_{(i)}$ occurred. It is to be noted that the expressions for E(N) obtained in [18] for several configurations of the cell-probabilities are directly valid here because it depends only on the sampling scheme and not the selection procedure used. For the configuration $\theta_1 = \ldots = \theta_k = \frac{1}{k}$, Panchapakesan [70] has shown that

(5.4.12) $$E(N) = M+M(k-1)k\frac{1}{B(M,\ldots,M,M-1)}\int_1^{\infty}\cdots\int_1^{\infty}\frac{(\prod_2^{k-1} y_i)^{M-1} y_1^{M-2}}{(1+\sum_1^{k-1} y_i)^{k(M+1)-2}}\prod_1^{k-1} dy_i .$$

In particular, when k = 2,

(5.4.13) $\sup_{\Omega} E(N) = M+2MI_{\frac{1}{2}}(M+1,M-1)$.

It is easy to show that when k = 2,

(5.4.14) $\sup_{\Omega} E(S|R_{14}) = 2I_{\frac{1}{2}}(M-D,M)$.

BIBLIOGRAPHY

1. Alam, K. (1971). Selection from Poisson processes. Ann. Inst. Statist. Math., 23, 411-418.

2. Alam, K., Seo, K., and Thompson, J. R. (1971). A sequential sampling rule for selecting the most probable multinomial event. Ann. Inst. Statist. Math., 23, 365-374.

3. Arimoto, S. (1971). Information-theoretical conservations on estimation problems. Information and Control, 19, 181-194.

4. Bahadur, R. R. (1950). On a problem in the theory of k populations. Ann. Math. Statist., 21, 362-375.

5. Bahadur, R. R., and Goodman, L. A. (1952). Impartial decision rules and sufficient statistics. Ann. Math. Statist., 23, 553-562.

6. Barlow, R. E., and Gupta, S. S. (1969). Selection procedures for restricted families of distributions. Ann. Math. Statist., 40, 905-917.

7. Barr, D. R., and Rizvi, M. H. (1966). An introduction to ranking and selection procedures. J. Amer. Statist. Assoc., 61, 640-646.

8. Barron, A. M., and Gupta, S. S. (1972). A class of non-eliminating sequential multiple decision procedures. Operation-Research Verfahren (ed. Henn, Künzi and Schubert), 11-37, Verlag Anton Hain Meisenheim am Glan, Germany.

9. Bartlett, N. S., and Govindarajulu, Z. (1965). Some distribution-free statistics and their application to the selection problem. Ann. Inst. Statist. Math., 20, 79-97.

10. Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

11. Bechhofer, R. E., Dunnett, C. W., and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. Biometrika, 41, 170-176.

12. Bechhofer, R. C., Elmarghraby, S., and Morse, N. (1959). A single-sample multiple decision procedure for selecting the multinomial event which has the largest probability. Ann. Math. Statist., 30, 102-119.

13. Bechhofer, R. E., Kiefer, J., and Sobel, M. (1968). Sequential Identification and Ranking Procedures. The University of Chicago Press, Chicago.

14. Bechhofer, R. E., Sobel, M. (1954). A single sample multiple decision procedure for ranking variances of normal populations. Ann. Math. Statist., 25, 273-289.

15. Bhapkar, V. P. and Gore, A. P. (1971). Some selection procedures based on U-statistics for the location and scale problems. Ann. Inst. Statist. Math., 23, 375-386.

16. Bickel, P. J. (1965). On some robust estimates of location. Ann. Math. Statist., 36, 847-858.

17. Brillouin, L. (1956). Science and Information Theory. Academic Press, New York.

18. Cacoullos, T., and Sobel, M. (1966). An inverse sampling procedure for selecting the most probable event in a multinomial distribution. Multivariate Analysis (ed. Krishnaiah, P. R.) 423-455, Academic Press, New York.

19. Chapman, D. G. (1952). On tests and estimates for the ratio of Poisson means. Ann. Inst. Statist. Math., 4, 45-49.

20. Csiszar, I. (1963). Eine informations theoretische Ungleichung und ihre anwendung auf deu beweis der ergodizität von markoffschen ketten. Publ. Math. Inst. Hungarian Acad. Sci., Ser. A., Fasc. 1-1, 8, 85-108.

21. DeGroot, M. H. (1970). Optimal Statistical Decisions. McGraw-Hill, New York.

22. Dudewicz, E. J. (1972). Statistical inference with unknown and unequal variances. Trans. Ann. Qual. Contr. Conf., Rochester Quality Control, 28, 71-85.

23. Dudewicz, E. J. (1974). A note on selection procedures with unequal observation numbers. Zastosowania Matematyki, 14, 31-35.

24. Eaton, M. L. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist., 38, 124-137.

25. Feinstein, A. (1958). Foundations of Information Theory. McGraw-Hill, New York.

26. Gnanadesikan, M. (1966). Some selection and ranking procedures for multivariate normal populations. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

27. Gnanadesikan, M., and Gupta, S. S. (1970). A selection procedure for multivariate normal distributions in terms of generalized variances. Technometrics, 12, 103-117.

28. Goel, P. K. (1972). A note on the non-existence of subset selection procedure for Poisson populations. Mimeo. Series No. 303, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

29. Goel, P. K. (1975). A note on subset selection procedure with Poisson processes. Unpublished report.

30. Goel, P. K., and Rubin, H. (1975). On selecting a subset containing the best population - A Bayesian approach. Mimeo Series No. 432, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

31. Govindarajulu, Z., and Harvey, C. (1974). Bayesian procedures for ranking and selection problems. Ann. Inst. Statist. Math., 26, 35-53.

32. Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo Series No. 150, Inst. of Statist., University of North Carolina, Chapel Hill, North Carolina.

33. Gupta. S. S. (1963). Probability integrals of multivariate normal and multivariate t. Ann. Math. Statist., 34, 792-828.

34. Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math., 14, 199-216.

35. Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics, 7, 225-245.

36. Gupta, S. S., and Deely, J. J. (1968). On the properties of subset selection procedures. Sankhyā, Ser. A, 30, 37-50.

37. Gupta, S. S., and Huang, D. Y. (1974a). Nonparametric subset selection procedures for the t-best populations. Bull. Inst. Math. Acad. Sinica, 2, 377-386.

38. Gupta, S. S., and Huang, D. Y. (1974b). On subset selection procedures for the entropy function associated with the binomial populations. To appear in Sankhyā.

39. Gupta, S. S., and Huang, D. Y. (1975). On subset selection procedures for Poisson populations and some applications to the multinomial selection problems. Applied Probability (ed. Gupta, R. P.) 97-109, North-Holland Publishing Co., Amsterdam.

40. Gupta, S. S., and Huang, D. Y. (1976). Selection procedures for the means and variances of normal populations when the samples sizes are unequal. To appear in Sankhyā.

41. Gupta, S. S., Huang, D. Y., and Huang, W. T. (1974). On a conditional procedure for selecting a subset containing the best of several binomial populations. Mimeo Series No. 340, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

42. Gupta, S. S., and Huang, W. T. (1974). A note on selecting a subset of normal population with unequal sample sizes. Sankhyā, Ser. A, 36, 389-396.

43. Gupta, S. S., and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (ed. Puri, M. L.) 419-514. Cambridge University Press, London.

44. Gupta, S. S., and Nagel, K. (1967). On selection and ranking procedures and order statistics from multinomial distributions. Sankhyā, Ser. B., 29, 1-34.

45. Gupta, S. S., Nagel, K., and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. Biometrika, 64, 403-413.

46. Gupta, S. S., and Panchapakesan, S. (1972). On multiple decision procedures. Ann. Math. Statist., J. Mathematical and Physical Sci., 6, 1-72.

47. Gupta, S. S., and Santner, T. J. (1973). On selection and ranking procedures - A restricted subset selection rule. Proceedings of 39th Session of the International Statistical Institute, Vienna, Austria, August 20-30, Contributed Papers, Vol. I, 409-417.

48. Gupta, S. S., and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to Probability and Statistics (ed. Olkin, I., at. el.), 224-248, Stanford University, Stanford, California.

49. Gupta, S. S., and Studden, W. J. (1970). On some selection and ranking procedures with applications to multivariate populations. Essays in Probability and Statistics (ed. Bose, R. C., et. al.)

50. Hardy, G. H., Littlewood, J. E., and Polyá, G. (1952). Inequalities. Cambridge University Press, London and New York.

51. Hettmansperger, T. P. (1968). On the trimmed Mann-Whitney statistic. Ann. Math. Statist., 39, 1610-1614.

52. Huang, D. Y. (1975). Some contributions to fixed sample and sequential multiple decision (selection and ranking) theory. Mimeo Series No. 363, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

53. Huang, W. T. (1972). Some contributions to sequential selection and ranking procedures. Mimeo Series No. 295, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

54. Huber, P. J. (1964). Robust estimation of a location parameter. Ann. Math. Statist., 35, 73-102.

55. Karlin, S. (1974). Inequalities for symmetric sampling plan I. Ann. Statist., 2, 1065-1094.

56. Kendall, D. G. (1963). Functional equations in information theory. Z. Wahrscheinlichkeitstheorie verw. Gebiete, 2, 225-229.

57. Khinchin, A. I. (1957). Mathematical Foundation of Information Theory. Dover Publication.

58. Kullback, S. (1959). Information Theory and Statistics, John Wiley, New York.

59. Lee, P. M. (1964). On the axioms of information theory, Ann. Math. Statist., 35, 414-441.

60. Lehmann, E. L. (1959). Testing Statistical Hypothesis, John Wiley, New York.

61. Lehmann, E. L. (1961). Some Model I problems of selection. Ann. Math. Statist., 32, 990-1012.

62. Lehmann, E. L. (1963). A class of selection procedures based on ranks. Math. Ann. 150, 268-275.

63. Lehmann, E. L. (1966). On a theorem of Bahadur and Goodman, Ann. Math. Statist., 37, 1-6.

64. Marshall, A. W., Olkin, I., and Proschan, F. (1967). Monotonicity of ratio of means and other applications of majorization. Inequalities (ed. Shisha, O.), 177-190, Academic Press, New York.

65. Mishina, A. P., and Proskuryakov, I. V. (1965). Higher Algebra, Pergamon Press, Oxford, London.

66. Mosteller, F. (1948). A k-sample slippage test for an extreme population. Ann. Math. Statist., 19, 58-65.

67. Nevius, S. E., Proschan, F., and Sethuraman, J. (1974). Schur functions in statistics III: A stochastic version of weak majorization with applications. Statistical Report No. M307, Department of Statistics, The Florida State University, Tallahassee, Florida.

68. Olkin, I., and Sobel, M. (1965). Integral expressions for tail probabilities of the multinomial and negative multinomial distributions. Biometrika, 52, 167-179.

69. Ostrowski, A. (1952). Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. J. Math. Pures Appl., 31, 303-307.

70. Panchapakesan, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. Statistical Decision Theory and Related Topics (ed. Gupta, S. S., and Yackel, J.), 275-298. Academic Press, New York.

71. Paulson, E. (1949). A multiple decision procedure for certain problem in analysis of variance. Ann. Math. Statist., 20, 95-98.

72. Perez, A. (1968). Risk estimates in terms of generalized f-entropies. Proceeding of the Colloquium on Information Theory (ed. Rényi, A.), 299-314, János Bolyai Mathematical Society, Budapest, Hungary.

73. Pfanzagl, J. (1974). On the Behrens-Fisher problem. Biometrika, 61, 39-47.

74. Puri, P. S., and Puri, M. L. (1967). Selection procedures based on ranks: scale parameter case. Sankhyā, Ser. A, 30, 291-302.

75. Quastler, H. (1953). Information Theory in Biology. University of Illinois Press, Urbana, Illinois.

76. Quastler, H. (1956). Information Theory in Psychology. Free Press, Glencoe, Illinois.

77. Rényi, A. (1960). On measures of entropy and information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 547-561, University of California Press, Berkeley and Los Angeles.

78. Rényi, A. (1967). On some basic problems of statistics: From point of view of information theory. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, 531-543, University of California Press, Berkeley and Los Angeles.

79. Rinott, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. Israel J. Math., 15, 60-77.

80. Rinott, Y. (1974). On two stage procedures for selecting the population with the largest mean from several normal populations with unknown variances. Unpublished Report.

81. Rizvi, M. H., and Sobel, M. (1967). Nonparametric procedures for selecting a subset containing the population with the largest α-quantile. Ann. Math. Statist., 38, 1788-1803.

82. Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, 131-148, University of California Press, Berkeley and Los Angeles.

83. Santner, T. (1975). A restricted subset selection approach to ranking and selection problems. Ann. Statist., 3, 334-349.

84. Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist., 26, 387-398.

85. Seal, K. C. (1957). An optimum decision rule for ranking means of normal populations. Calcutta Statist. Assoc. Bull., 7, 131-150.

86. Seal, K. C. (1958). On ranking parameters of scale in type III populations. J. Amer. Statist. Assoc., 53, 164-175.

87. Shannon, C. E. (1948). A mathematical theory of communication. Bell System Tech. J. 27, 373-423.

88. Sitek, M. (1972). Application of the selection procedure R to unequal observation numbers. Zastosowania Matematyki, 12, 355-371.

89. Sobel, M. (1963). Single sample ranking problems with Poisson populations. Tech. Report No. 19, Department of Statistics, University of Minnesota, Minneapolis, Minnesota.

90. Sobel, M., and Huyett, M. H. (1957). Selecting the best one of several binomial populations. Bell System Tech. J. 36, 537-576.

91. Sobel, M., and Weiss, G. H. (1972). Play-the-winner rule and inverse sampling for selecting the best of $k \geq 3$ binomial populations. Ann. Math. Statist., 43, 1808-1826.

92. Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist., 16, 243-258.

93. Studden, W. J. (1967). On selecting a subset of k populations containing the best. Ann. Math. Statist., 38, 1072-1078.

94. Sugiura, N. (1965). Multisample and multivariate nonparametric tests based U-statistics and their asymptotic efficiencies. Osaka J. Math., 2, 385-426.

95. Tamura, R. (1971). On a c-sample test based on trimmed samples. Ann. Math. Statist., 42, 1455-1460.

96. Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics (ed. Olkin, I. et. al.) 448-4 85, Stanford University Press, Stanford, California.

97. Vajda, I. (1969). A contribution to the informational analysis of pattern. Methodologies of Pattern Recognition, 509-519, Academic Press, New York.

98. Vincze, I. (1968). On the information-theoretical foundation of mathematical statistics, Proceedings of the Colloquium on Information Theory (ed. Rényi, A.), 503-509, János Bolyai Mathematical Society, Budapest, Hungary.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Mimeograph Series #454

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Some contributions to subset selection procedures

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER
Mimeo. Series #454

7. AUTHOR(s)
Wing-Yue Wong

8. CONTRACT OR GRANT NUMBER(s)
N00014-75-C-0455

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
Purdue University
W. Lafayette, IN 47907

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
0455

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Washington, DC

12. REPORT DATE
May, 1976

13. NUMBER OF PAGES
118

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Subset selection procedure, Normal population, U-Statistics, Poisson processes, Finite scheme, Majorization, Schur functions, Negative multinomial distribution, Inverse sampling rule.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
The main purpose of this paper is to propose and study the subset selection approach for some new problems and make contributions. Chapter I deals with some selection and ranking procedures for the largest unknown mean of k normal populations with unequal variances. The procedures are based on unequal number of observations from the given k normal populations. An indifference zone approach to the problem of selecting the populations with the t-largest unknown means is also studied. In Chapter II some nonparametric subset selection procedures based on U-statistics for selecting the largest of the k location

parameters are proposed and studied. Again, the procedures are based on unequal number of observations from each of the k populations. Chapter III discusses some subset selection procedures for Poisson processes. Some properties of the proposed selection rules are discussed. Applications to binomial and multinomial selection problems are given. Chapter IV deals with a class of selection rules for finite schemes. The parameter space is partially ordered by means of majorization and the proposed selection rules are based on Schur functions. An application to testing the homogeneity of k finite schemes is given. Chapter V discusses some subset selection procedures for a negative multinomial distribution. An inverse sampling rule for selecting the cell with largest cell-probability from a multinomial distribution is considered.